

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 16, 2003, 08:09:05 ; Search time 4508 seconds
(without alignments)

10684.371 Million cell updates/sec

Title: US-09-700-696b-1

Perfect score: 1655
Sequence: 1 gtgataaagataatagat.....ataaaaaaaaaaaaaaaaaa 1655

Scoring table: IDENTITY_NUC
Gapop 10.0 ; Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl: *
1: gb_da: *
2: gb_hlg: *
3: gb_in: *
4: gb_cm: *
5: gb_ov: *
6: gb_ph: *
7: gb_pl: *
8: gb_pr: *
9: gb_ro: *
10: gb_sts: *
11: gb_sy: *
12: gb_un: *
13: gb_vl: *
14: em_da: *
15: em_fun: *
16: em_hum: *
17: em_mu: *
18: em_mu: *
19: em_mu: *
20: em_mu: *
21: em_mu: *
22: em_mu: *
23: em_mu: *
24: em_mu: *
25: em_mu: *
26: em_mu: *
27: em_mu: *
28: em_mu: *
29: em_mu: *
30: em_mu: *
31: em_mu: *
32: em_mu: *
33: em_mu: *
34: em_mu: *
35: em_mu: *
36: em_mu: *
37: em_mu: *
38: em_mu: *
39: em_mu: *
40: em_mu: *
41: em_mu: *

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1655	100.0	1655	6	AX010293	AX010293 Sequence
2	1655	100.0	1655	6	AX135682	AX135682 Sequence
3	1654	99.9	1989	6	HS4276396	AX135682 Homo sapi
4	1654	99.9	2013	6	AX135707	AX135707 Sequence
5	1638	99.0	187624	9	AC093768	AC093768 Homo sapi
6	1510.4	91.3	2095	9	AB046056	AB046056 Macaca fa
7	1508.4	91.1	2140	9	AB050259	AB050259 Macaca fa
8	1507.6	91.1	2259	9	AB060891	AB060891 Macaca fa
9	1496.2	90.4	2085	9	AB056814	AB056814 Macaca fa
10	1292	78.1	1617	9	AF325916	AF325916 Homo sapi
11	532.6	32.2	1651	10	AF330559	AF330559 Rattus no
12	532.6	32.2	1672	10	AF330558	AF330558 Rattus no
13	522	31.5	169603	2	AC129695	AC129695 Rattus no
14	514.8	29.3	1655	10	AF260922	AF260922 Rattus no
15	485.2	29.3	1682	10	AF298661	AF298661 Mus muscu
16	472.2	28.5	1580	10	AF314964	AF314964 Mus muscu
17	245.4	14.8	6058	2	AC124106	AC124106 Mus muscu
18	80.4	4.9	68001	2	AC122775	AC122775 Mus muscu
19	77	4.7	60588	2	AC124106	AC124106 Mus muscu
20	70	4.2	7218	6	166494	AC124106 Mus muscu
21	59.2	3.6	172853	9	AC084361	AC084361 Homo sapi
22	55	3.3	91048	9	AC007926	AC007926 Homo sapi
23	54.6	3.3	135204	9	AC007926	AC007926 Homo sapi
24	51.6	3.1	115758	9	AC104634	AC104634 Homo sapi
25	51.4	3.1	184535	2	CNS05TCU	AL355100 Homo sapi
26	50.8	3.1	5586	6	AX348391	AX348391 Sequence
27	50.4	3.0	124820	2	AC117073	AC117073 Dictyoste
28	49.8	3.0	81120	2	AC022851	AC022851 Homo sapi
29	49.8	3.0	268147	2	AC116965	AC116965 Dictyoste
30	49.2	3.0	10176	3	AE001370	AE001370 Plasmodu
31	49.2	3.0	43993	2	AC116965	AC116965 Dictyoste
32	49.2	3.0	214229	9	CNS08CNA	AC116965 Dictyoste
33	49	3.0	132914	2	AE001383	AE001383 Plasmodu
34	48.8	2.9	81179	2	AC116100	AC116100 Dictyoste
35	48.6	2.9	14459	3	AE001383	AE001383 Plasmodu
36	48.6	2.9	100726	2	AC116961	AC116961 Dictyoste
37	48.6	2.9	169918	2	AC129254	AC129254 Rattus no
38	48.4	2.9	3801	3	PFDAECPN	AC129254 Rattus no
39	48.2	2.9	77096	9	AC116030	AC116030 Dictyoste
40	48.2	2.9	102794	2	AC104623	AC104623 Homo sapi
41	48.2	2.9	129360	2	AC117079	AC117079 Dictyoste
42	48.2	2.9	146383	2	AC116367	AC116367 Oryza sat
43	48.2	2.9	192929	2	AC005505	AC005505 Plasmodu
44	48	2.9	2277	6	AR004980	AR004980 Sequence
45	47.8	2.9	110236	2	AC116032	AC116032 Dictyoste

ALIGNMENTS

RESULT 1	AX010293	1655 bp	DNA	Linear	PAT 06-SEP-2000
LOCUS	AX010293	Sequence 1 from Patent WO960017.			
DEFINITION	AX010293				
ACCESSION	AX010293.1	GI:9997173			
VERSION					
KEYWORDS					
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Rowe, P.	A novel polypeptide hormone phosphatonin			
AUTHORS	1 (bases 1 to 1655)				
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.				
JOURNAL	Patent: WO 960017-A 1 25-NOV-1999;				

Pred. No. is the number of results predicted by chance to have a

REFERENCE	AUTHORS	TITLE	JOURNAL	FEATURES	BASE COUNT	ORIGIN
ORGANISM	Homio sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;					
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.					
TITLE	1 (bases 1 to 1655)					
JOURNAL	Rowe, P.					
FEATURES	Polypeptide hormone phosphatonin					
FEATURES	Patent: WO 0132878-A 1 10-MAY-2001;					
FEATURES	UNIVERSITY COLLEGE LONDON (GB)					
FEATURES	Location/Qualifiers					
FEATURES	1. 1655					
FEATURES	/organism="Homo sapiens"					
FEATURES	/db_xref="taxon:9606"					
BASE COUNT	609 a	303 c	380 g	363 t		
ORIGIN						
Query Match	100.0%;	Score 1655;	DB 6;	Length 1655;		
Best Local Similarity	100.0%;	Pred. No. 0;				
Matches 1655:	Conservative	0;	Mismatches	0;	Indels	0;
						Gaps 0;
1	GTGATTAAGAATTAATAGATCAGTAAACAAAGAGATTAATCTCACAATGGCCCTGAGATGTCA	60				
1	GTGATTAAGAATTAATAGATCAGTAAACAAAGAGATTAATCTCACAATGGCCCTGAGATGTCA	60				
61	ATTATTCCTAAGTCACTGGGAAATAAGGGTTGAGATGAGATGATCTATCAGCAAA	120				
61	ATTATTCCTAAGTCACTGGGAAATAAGGGTTGAGATGAGATGATCTATCAGCAAA	120				
121	CTACATGACCAAGAAATATGGCCAGCTCTCATCAGAAATTAACATGCAACATATATATG	180				
121	CTACATGACCAAGAAATATGGCCAGCTCTCATCAGAAATTAACATGCAACATATATATG	180				
181	GGGCCAGTACTGCGATTAACTCCTGGGGGAGAGAAACAAAGAGAACACACTAGAAAT	240				
181	GGGCCAGTACTGCGATTAACTCCTGGGGGAGAGAAACAAAGAGAACACACTAGAAAT	240				
241	GTTCCTAACATTAATCCACAGATATGATTTATGCTAAGACACATCGAAGATTAATG	300				
241	GTTCCTAACATTAATCCACAGATATGATTTATGCTAAGACACATCGAAGATTAATG	300				
301	AAGCCTCAAGAGATATCCCAAGCCAGAAAGATCCAGTAAAGAGCAAGCAAGCCATCGT	360				
301	AAGCCTCAAGAGATATCCCAAGCCAGAAAGATCCAGTAAAGAGCAAGCAAGCCATCGT	360				
361	ATTCAACACACATTTGACTACTTAATAACATCTCTCAAAAATCCCAAGTAT	420				
361	ATTCAACACACATTTGACTACTTAATAACATCTCTCAAAAATCCCAAGTAT	420				
421	TTTGAAGGAGGGGTTATACAGATCTTCAAGAGAGAGGGGACAAATATATCTCTTC	480				
421	TTTGAAGGAGGGGTTATACAGATCTTCAAGAGAGAGGGGACAAATATATCTCTTC	480				
481	AGTGGGAGCGGCAACCTTTTAAGACATCTCTGTAAAGAGAACTACTGGTCTGAC	540				
481	AGTGGGAGCGGCAACCTTTTAAGACATCTCTGTAAAGAGAACTACTGGTCTGAC	540				
541	CTAGAAAGGCAAAAGATTTCAAAACAGGGTTTGCAGGCCCAAGTAAGTGAATCTCAT	600				
541	CTAGAAAGGCAAAAGATTTCAAAACAGGGTTTGCAGGCCCAAGTAAGTGAATCTCAT	600				
601	CTTGACACAAAAGAGCCAGTTATATATGAGATCCAGAGAGAGAAATGGTGGAAAT	660				
601	CTTGACACAAAAGAGCCAGTTATATATGAGATCCAGAGAGAGAAATGGTGGAAAT	660				
661	ACCATTTGGCAATAGGATATAACTGCGAAAGAGGCGAGATCTGTATGTCAAGCTTTGTA	720				
661	ACCATTTGGCAATAGGATATAACTGCGAAAGAGGCGAGATCTGTATGTCAAGCTTTGTA	720				
721	GAGGCGAGCAAGATATCAATGGGTACTACCAATTTTAAGGAGCTCCCTGGAGAGAGGA	780				
721	GAGGCGAGCAAGATATCAATGGGTACTACCAATTTTAAGGAGCTCCCTGGAGAGAGGA	780				
781	AACAGAGTGAATGCTGGCAGCCAAAATGCTCACCAAGGAGGTTAGTTCAATACCT	840				

Db	781	AAAGAGGAGTGCATGCGAGCGCAAAATGCTCACCAGGGAAGGTTGAGTTTCATTACCT	840
Qy	841	CCTGCACCCCTCAAAAGAGAAAAGAAAGAGCAGCTAGTATGCGAGCTGAAAAGTACCAAC	900
Db	841	CCTGCACCCCTCAAAAGAGAAAAGAAAGAGCAGCTAGTATGCGAGCTGAAAAGTACCAAC	900
Qy	901	TATATGAAATTCCTAAAAATGCGAAAGCGATGACAGAAAGGGGTGATCATCTAT	960
Db	901	TATATGAAATTCCTAAAAATGCGAAAGCGATGACAGAAAGGGGTGATCATCTAT	960
Qy	961	AGGAACCAAGCACCCTAAATGAAAAACAAGGTTCTTAGTAGGGCAAAAGTCAGGGC	1020
Db	961	AGGAACCAAGCACCCTAAATGAAAAACAAGGTTCTTAGTAGGGCAAAAGTCAGGGC	1020
Qy	1021	CTGGCCATTCCTTCGCGGCTTGATATATGAAATCAAAAACGAAATGATTCCTTAT	1080
Db	1021	CTGGCCATTCCTTCGCGGCTTGATATATGAAATCAAAAACGAAATGATTCCTTAT	1080
Qy	1081	GGCCCCAGTCATGAGAAATATATAACACATGGCAGAAAATATCATTTATGCCACAGA	1140
Db	1081	GGCCCCAGTCATGAGAAATATATAACACATGGCAGAAAATATCATTTATGCCACAGA	1140
Qy	1141	CAAAATATTTCTACACGGAATTAAGGTATGCCACAAGGAAAGCTCCTGGGGTAGACA	1200
Db	1141	CAAAATATTTCTACACGGAATTAAGGTATGCCACAAGGAAAGCTCCTGGGGTAGACA	1200
Qy	1201	CCCCATTCACAGAGAGTTTAGTTCCCGTAGAAGAGATGACAGTATGATCATCTGAC	1260
Db	1201	CCCCATTCACAGAGAGTTTAGTTCCCGTAGAAGAGATGACAGTATGATCATCTGAC	1260
Qy	1261	AGTGCAGTTCAAGTAGAGACGATGTGACTAGTCCACAGAGATTCACAGCGGGGTGAC	1320
Db	1261	AGTGCAGTTCAAGTAGAGACGATGTGACTAGTCCACAGAGATTCACAGCGGGGTGAC	1320
Qy	1321	AGTGCAGACCTCGTACACCTGTGATGATGATAGAGAGACCACTGACAGTACCA	1380
Db	1321	AGTGCAGACCTCGTACACCTGTGATGATGATAGAGAGACCACTGACAGTACCA	1380
Qy	1381	GGTGAAGAGAGATAGAGTGAAGAACTGAGTAGGCCAAGAAATCGTCTCCCTGGGGGA	1440
Db	1381	GGTGAAGAGAGATAGAGTGAAGAACTGAGTAGGCCAAGAAATCGTCTCCCTGGGGGA	1440
Qy	1441	ATTTTTCATCTTATATATCTACAGTATTAATTTCTATTAAGGCTATTAAGTTTAAAG	1500
Db	1441	ATTTTTCATCTTATATATCTACAGTATTAATTTCTATTAAGGCTATTAAGTTTAAAG	1500
Qy	1501	CAAAAAAATCATTAACAGATCTATGAATAGTAACATTTAGTAGGTGTCATTAAAA	1560
Db	1501	CAAAAAAATCATTAACAGATCTATGAATAGTAACATTTAGTAGGTGTCATTAAAA	1560
Qy	1561	ATAGTTGGTAATGTCACAAATGCCCTTCTATGTTGTTGCTCTGTAAGACATGAAAAT	1620
Db	1561	ATAGTTGGTAATGTCACAAATGCCCTTCTATGTTGTTGCTCTGTAAGACATGAAAAT	1620
Qy	1621	CAATATCTCGATGATATAAAAAAATAAAAA	1655
Db	1621	CAATATCTCGATGATATAAAAAAATAAAAA	1655
RESULT 3			
HSA276396			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			

Center: Washington University Genome Sequencing Center
 Center code: W06SC
 Web site: <http://genome.wustl.edu/gsc>
 Contact: sapiens@wustl.wustl.edu
 ----- Summary Statistics -----
 Center project name: H.NH0113613
 Drafting Center: WIBR

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

SOURCE INFORMATION: The RPEC-11 human BAC library was made from the blood of one male donor, as described by Oseguwa, K., Moon, P. Y., Zhao, B., Frengen, E., Tatemio, M., Catanesi, J. J., and de Jong, P. J. (1998). An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the right is AC087106. Actual start of this clone is at base position 1 of Rp11-113613; actual end is at base position 187624 of Rp11-113613.

Sequence derived from one plasmid subclone, base position 69516 to 69532.

The sequence of AC021959 has been incorporated into AC093768.

FEATURES	location/Qualifiers
source	1..187624
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="4"
	/map="4"
	/clone="RP11-113G13"
	/clone_lib="RPC1-11"
	6069..6195
repeat_region	/rpt_family="L1"
	6288..6308
repeat_region	/rpt_family="AT_rich"
	6422..6530
repeat_region	/rpt_family="L1"
	6559..6582
repeat_region	/rpt_family="AT_rich"
	6708..6744
repeat_region	/rpt_family="AT_rich"
	6732..7028
repeat_region	/rpt_family="Alu"
	8033..8053
repeat_region	/rpt_family="(TTTAA)n"
	8041..8459
repeat_region	/rpt_family="L1"
	8904..8942
repeat_region	/rpt_family="MIR"
	8943..9214
repeat_region	/rpt_family="MaLR"
	9526..9591
repeat_region	/rpt_family="(TA)n"
	10055..10530
repeat_region	/rpt_family="L1"
	11326..11417
repeat_region	/rpt_family="L2"
	11399..11445
repeat_region	/rpt_family="(TGA)n"
	15645..15865
repeat_region	/rpt_family="L2"
	15933..16227

misc-feature /rpt-family="Alu" .16161
 complement(16160. .16161)
 /note="similar to Homo sapiens EST A1745652 (NID:95113940)
 tr24c10.x1"
 16226. .16251
 repeat-region /rpt-family="(CMA)n"
 16337. .16436
 misc-feature /note="similar to Homo sapiens EST AA873157 (NID:92969279)
 oh73g11.s1"
 16340. .16915
 misc-feature /note="similar to Homo sapiens EST AV702108
 (NID:910718438)"
 16340. .16899
 misc-feature /note="similar to Sus scrofa EST B1181092 (NID:914655501)"
 16340. .16447
 misc-feature /note="similar to Rattus norvegicus EST BF391035
 (NID:911375882)"
 16340. .16436
 misc-feature /note="similar to Homo sapiens EST AA651669 (NID:92583321)
 ns89604.s1"
 16340. .16436
 misc-feature /note="similar to Homo sapiens EST AA885538 (NID:93000646)
 n252802.s1"
 16340. .16436
 misc-feature /note="similar to Homo sapiens EST A1088133 (NID:93427109)
 o295g07.x1"
 16340. .16436
 misc-feature /note="similar to Homo sapiens EST AL533242
 (NID:912796735)"
 16340. .16436
 misc-feature /note="similar to Homo sapiens EST BE742618
 (NID:910156597)"
 16340. .16436
 misc-feature /note="similar to Homo sapiens EST BI094057
 (NID:914512387)"
 16342. .16436
 misc-feature /note="similar to Homo sapiens EST AA398444 (NID:92051756)
 zt74g07.s1"
 16355. .16380
 misc-feature /note="similar to Homo sapiens EST W00667 (NID:91272102)
 za72b10.r1"
 16388. .16436
 misc-feature /note="similar to Homo sapiens EST BE621830
 (NID:99892770)"
 16403. .17080
 misc-feature /note="similar to Homo sapiens EST BI084966
 (NID:914503296)"
 16405. .16436
 misc-feature /note="similar to Homo sapiens EST BI522324
 (NID:915347116)"
 16406. .16436
 misc-feature /note="similar to Homo sapiens EST A1745652 (NID:95113940)
 tr24c10.x1"
 16458. .17080
 misc-feature /note="similar to Homo sapiens EST AL533242
 (NID:912796735)"
 16458. .17080
 misc-feature /note="similar to Homo sapiens EST BE621830
 (NID:99892770)"
 16458. .17080
 misc-feature /note="similar to Homo sapiens EST BE742618
 (NID:910156597)"
 16458. .17077
 misc-feature /note="similar to Homo sapiens EST BI522324
 (NID:915347116)"
 16458. .16873
 misc-feature /note="similar to Homo sapiens EST AA398444 (NID:92051756)
 zt74g07.s1"
 16458. .16868
 misc-feature /note="similar to Homo sapiens EST BI094057
 (NID:910156597)"

Query Match 99.0%; Score 1638; DB 9; Length 187624;
 Best Local Similarity 100.0%; Pred. No. 0;

Matches 1638; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 2 TGAATTAAGATATAGTATACGTAACAAAGAGATATCTCAATGGCTGAGATGCA 61
 75854 TGAATTAAGATATAGTATACGTAACAAAGAGATATCTCAATGGCTGAGATGCA 75913
 62 TTTATCTTAAGTCACTGGGATTAAGGGTTTGGATGAGATGATGATACGCAAC 121
 75914 TTTATCTTAAGTCACTGGGATTAAGGGTTTGGATGAGATGATGATACGCAAC 75973
 122 TACATGACCAAGAAATATGCGCAGCTCTCATACAAATTAATCAATCAATTAATG 181
 75974 TACATGACCAAGAAATATGCGCAGCTCTCATACAAATTAATCAATCAATTAATG 76033
 182 GGCACGTACGTGATTAATCTCTGGGGGAAAGAAACAAAGAGACACCTAGAAATG 241
 76034 GGCACGTACGTGATTAATCTCTGGGGGAAAGAAACAAAGAGACACCTAGAAATG 76093
 242 TTCTTAACATTAATCCCAAGATATGAATTTATGCTTAAAGCAGCTCGAAGGATTAAGA 301
 76094 TTCTTAACATTAATCCCAAGATATGAATTTATGCTTAAAGCAGCTCGAAGGATTAAGA 76153
 302 AGCTCAAGAGATTTCCCAAGCCCAAGAAAGTCCAGATTAAGAAAGCAAGCCATGGTA 361
 76154 AGCTCAAGAGATTTCCCAAGCCCAAGAAAGTCCAGATTAAGAAAGCAAGCCATGGTA 76213
 362 TTCAACACAAATGACTTACCTTAACATCTCTCTCAAAAGTCAAAATCCCATGATTT 421
 76214 TTCAACACAAATGACTTACCTTAACATCTCTCTCAAAAGTCAAAATCCCATGATTT 76273
 422 TTGAAGCAGCGGTTATACAGATCTTCAAGAGAGAGGGGACATGATATCTCTTTCA 481
 76274 TTGAAGCAGCGGTTATACAGATCTTCAAGAGAGAGGGGACATGATATCTCTTTCA 76333
 482 GTGGGAGCGCCCAACCTTTTAAGGACATTTCTGTTAAAGGAGAACTCTGCTGACC 541
 76334 GTGGGAGCGCCCAACCTTTTAAGGACATTTCTGTTAAAGGAGAACTCTGCTGACC 76393
 542 TGAAGGCAAGATATTTCAACAGAGTTTGCAGGCCCAAGTGAAGTACTATCTC 601
 76394 TGAAGGCAAGATATTTCAACAGAGTTTGCAGGCCCAAGTGAAGTACTATCTC 76453
 602 TTGACACAAAAAGCCAGGTTATATGAGATCTCCAGAGAGAGAGAAATGTTGAATA 661
 76454 TTGACACAAAAAGCCAGGTTATATGAGATCTCCAGAGAGAGAGAAATGTTGAATA 76513
 662 CCATTGGAATCTAGGATGAAGTGAAGTGAAGGAGAGATCTGTTGATGTCAGCTTAG 721
 76514 CCATTGGAATCTAGGATGAAGTGAAGTGAAGGAGAGATCTGTTGATGTCAGCTTAG 76573
 722 AGGCGACGAACGATATCATGGGTAGTACCAATTTTAAGGAGCTCCCTGGAAGAGAGAA 781
 76574 AGGCGACGAACGATATCATGGGTAGTACCAATTTTAAGGAGCTCCCTGGAAGAGAGAA 76633
 782 ACAAGTGTATGCTGCGACGCAAAATGCTCACCAGAGGAAAGTTGATTTCAATCCCTC 841
 76634 ACAAGTGTATGCTGCGACGCAAAATGCTCACCAGAGGAAAGTTGATTTCAATCCCTC 76693
 842 CTGACCGCTCAAAAGAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 901
 76694 CTGACCGCTCAAAAGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 76753
 902 ATATGAAATTTCTTAATAATGCGCAAGGAGTACCAAGAAAGGTTGATGATTTCAATA 961
 76754 ATATGAAATTTCTTAATAATGCGCAAGGAGTACCAAGAAAGGTTGATGATTTCAATA 76813
 962 GGAACCAAGACCTTAATAAGAAACAAAGTTTCTTAAGGCAAGGTTGATGATTTCAATA 1021
 76814 GGAACCAAGACCTTAATAAGAAACAAAGTTTCTTAAGGCAAGGTTGATGATTTCAATA 76873
 1022 TGCCCATTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1081
 76874 TGCCCATTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 76933

RESULT 6	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
AB046056		2095 bp mRNA linear	AB046056	AB046056.1 GI:9280167	Macaca fascicularis adult cDNA to mRNA, clone-lib:macaque brain cDNA library QCEE clone:QCEE-12450.	Macaca fascicularis	Macaca fascicularis	Osada, N., Hida, M., Kusuda, J., Tanuma, R., Iseki, K., Hirai, M., Terao, K., Suzuki, Y., Sugano, S. and Hashimoto, K.	Isolation of full-length cDNA clones from macaque brain cDNA libraries	Submitted 2 (bases 1 to 2095)	Hashimoto, K., Osada, N., Hida, M., Kusuda, J. and Sugano, S.	Direct Submission	Submitted (14-JUL-2000)	Katsuyuki Hashimoto, National Institute of Infectious Diseases, Division of Genetic Resources; 23-1, Toyama 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan	(E-mail: khashi@nih.go.jp, url: http://www.nih.go.jp/yoken/genedbank/, Tel: 81-3-5285-1111 (ex 2150), Fax: 81-3-5285-1181)	URL: http://www.nih.go.jp/yoken/genedbank/ Lib Name: macaque brain cDNA library QCEE Lab host: TOP10

```

/product="matrix extracellular phosphoglycoprotein"
/protein_id="BAB01638.1"
/db_xref="GI:9280168"
/translation="MRVGVGLLELSTVAAPFPQPTERTKTSQCEBQRIITYKGNH
KHGHVFKVCVYMSPGKKNQDTQKQEKKNKDIMGHLGRKRYOELSKENIVQERK
LISLEASENNSSKSNQFYTRNRLNKYTSLSNKENINGLMSTIYPTSTNKPQAD
DDALSTLHDOEYEGALIRNNMHMGVTAIKLLGEKNQSPKNVNTKIPANMYA
KAKSKDKKPPORDSOVOYKVYKSKSTHTQNTQINIDYPRILSKYVKIKIPDFEESGTTDLC
ERGNMDSPEFGDQOPPRKDIPGKEGATSLEGDQIDQIGFAGSPSEASTNLDTKEG
NEILPERKENGNTIGTGETAEKADAVDSLVEGNQDQSTNGKLELGRGNREVG
GQNAHQGVFEFHYPPAPSKERKEKGSASTSTYNLEPRKDNQSAKRGVQDSNRNOA
ILHKKRQPSKSGKSGGLPIPSRGLDNETKMTDLSNGSNINIIHRSKRIHYVPRQNN
PTRKKGMFHGKSGWRDPSYNRRLSSRRRDESSSDSGSSSESDG"

```

Query Match Similarity	91.3%	Score 1510.4	DB 9	Length 2095
Best local Similarity	95.5%	Pred No. 0		
Matches 1577	Conservative 0	Mismatches 71	Indels 4	Gaps 2
Qy	2	TGAATTAAGAAATATAGTATCAGTACGTAACAAAGAGAAATCTACATGAGCGCTGAGATGTCAA	61	
Db	448	TGAATTAAGAAATATAGTATCAGTACGTAACAAAGAGAAATCTACATGAGCGCTGAGATGTCAA	507	
Qy	62	TTTATTCCTAAGTCACACTGGGAATTAAGGGTTTGAGAGTGGAGATGATGCTATCAGCAAC	121	
Db	508	TTTATTCCTAAGTCACACTGGGAATTAAGGGTTTGAGAGTGGAGATGATGCTATCAGCAAC	567	
Qy	122	TACATGACCAAGAAAGAAATTTGGCGAGCTCTCATGAGAAATTAACATGCAACATTAATATGG	181	
Db	568	TACATGACCAAGAAAGAAATTTGGCGAGCTCTCATGAGAAATTAACATGCAACATTAATATGG	627	
Qy	182	GGCCAGTACTGCGATTTAAACTCTGGGGGGAAGAAAACAAAGAGAACACACCTAGGAATG	241	
Db	628	GGCCAGTACTGCGATTTAAACTCTGGGGGGAAGAAAACAAAGAGAACACACCTAGGAATG	687	
Qy	242	TTTCTAATCAATATATCCAGCAGATATGAATTTATGCTTAAAGCACACTGGAAGATTAAAAAGA	301	
Db	688	TTTCTAATCAATATATCCAGCAGATATGAATTTATGCTTAAAGCACACTGGAAGATTAAAAAGA	747	
Qy	302	AGGCTCAAGAGAAATTTCCCAAGGCCGAGAAAGCCGATTAAGAAACAAAGACCCATCTCTTA	361	
Db	748	AGGCTCAAGAGAAATTTCCCAAGGCCGAGAAAGCCGATTAAGAAACAAAGACCCATCTCTTA	807	
Qy	362	TTTCAACACACAACTTACTACTCTAAACATCTCTCAAAAGTCAAAAAATATCCCATGATTT	421	
Db	808	TTTCAACACACAACTTACTACTCTAAACATCTCTCAAAAGTCAAAAAATATCCCATGATTT	867	
Qy	422	TTTGAAGCAGCGGTTATACAGATCTTCAAGAGAGAGGGGACAAATGATATATCTCTTTTCA	481	

Db	868	TTGAAAGCGACGGTTACACAGATCTTCAAGAGAGAGGGGCAAAATGATATGTCTCTTCA	927
OY	482	GTGGGGAGCGCAACCTTTTAAGACATTTCTCGTAAAGAGAGTACTGTCTGAC	541
Db	928	GTGGGAGCGGCAACCTTTTAAGGACATTTCTGGTAAAGGAGAAAGCTACTGTCTGAC	987
OY	542	TAGAAAGCAAGATATTCACACAGGTTTGCAGGCCCAAGTGAAGCTGAGAGTACATC	601
Db	988	TGAAAGCGAAGATTTTCAACACAGGTTTGCAGGCCCAAGTGAAGCTGAGAGTACTAATC	104
OY	602	TTGACACAAAAAGCCAGGTTATTAATGATCCAGAGAGAGAAATGTGGAATA	661
Db	1048	TTGACACAAAAAGCGAGGTTATTAATGATCCAGAGAGAGAAATGTGCGAATA	110
OY	662	CCATTGGAACTAGGATGAACCTCGGAAAGGCGAGATGCTGTGATGTACGCTTGTAG	721
Db	1108	CCATTGGAACTGGGATGAACCTCGGAAAGGCGAGATGCTGTGATGTACGCTTGTAG	116
OY	722	AGGCGAGCAGGATATCATGGGTAGTACATTTTAAGAGACTCCCTGGAAGAGAGAA	781
Db	1168	AGGCGAACAAGATATCATGGGTAGTACATTTTAAGAGACTCCCTGGAAGAGAGAA	122
OY	782	ACAGATGGATTCGTGGCAGCCAAATGCTCACCAAGGAAAGTTGATTCATTACCTC	841
Db	1228	ACAGATGGATTCGTGGCAGCCAAATGCTCACCAAGGAAAGTTGATTCATTACCTC	128
OY	842	CTGCACCTCTCAAAAGAGAAAAAGAAAGGACAGTAGTGTAGAGCTCCGTAAGAGAACT	901
Db	1288	CTGCACCTCTCAAAAGAGAAAAAGAAAGGACAGTAGTGTAGAGCTCCGTAAGAAAT	134
OY	902	ATAATGAATTCCTTAAANATGGCAAGGCACTACAGAAAGGGTGTAGTCAATCTTAATA	961
Db	1348	ATAATGAATTCCTTAAANATGGCAAGGCACTACAGAAAGGGTGTAGTCAATCTTAATA	1407
OY	962	GGAACCAAGCAACTTAAATGAAGAAAAAGAAAGTTTCTGTAGGGCAAAAGCAGAGGC	1021
Db	1408	GGAACCAAGCAATCTTACATGAAGAAAAAGAAAGTTTCTGTAGGGCAAAAGCAGAGGC	146
OY	1022	TGCCATTCTTTCGTGCTGTGGATATGAATAACAAAACGAATGATTCCTTTAAG	1081
Db	1468	TGCCATTCTTTCGTGCTGTGGATATGAATAACAAAACGAATGATTCCTTTAAG	1527
OY	1082	GCCCCAGTCAGAGATATTAATAACATGGGAGAAAAATATCTTATGTACCCACAGAC	1144
Db	1528	GCCCCAGTATGAGAAAT---ATACACACAGCAGAAAAATATCTTATGTACCCATAGAC	158
OY	1142	AAAAATATCTCACGGAATTAAGGATATGCCACAAAGGAAAGGCTCTGGGTAAACAAC	1201
Db	1585	AAAAATATCTCACGGAATTAAGGATATGCCACATGGAAGAAAGGCTCTGGGTAAACAAC	1644
OY	1202	CCCATTCACACAGGATTTAAGTTCCCGTGAAGAGGATGACAGTACGTATCTGACA	1261
Db	1645	CATTTCTCACACAGGATTTAAGTTCCCGTGAAGAGGAGACAGTATGATGTATCTGACA	1704
OY	1262	GTGGCAGTTCAAGTGAAGAGCGATGTGACTAGTCCACAGAGATTTCCAGCGGGGTGACA	1321
Db	1705	GTGGCAGTTCAAGTGAAGAGCGATGTGACTAGTCCACAGAGATTTCCAGCGGGGTGAC	1766
OY	1322	GTCGAGACCTCGTCAACCTGTAGATTGATTAAGAGAGAGCCACCTGACAGCTGACCAG	1381
Db	1765	GTCGAGAACCTCATCACTGTAGATTGACGTACAGAGAGAGCCACCTGACAGCTGACCAG	1824
OY	1382	GTGAAGAGAGATAGAGTGAAGAACTGAGTGAAGCCCAAGAAATCTGTCTCTGGGGGAA	1441
Db	1825	GTGAAGAGAGAGAGAGGTGAAGAAATGAGTGAAGCCCAAGAAATCTGTCTCTGGGGGAA	1884
OY	1442	TTTTTGTATCTTAATATGTACAGTATTAATAATCTATTAAGAGCTATATGTTTTTAAGC	1501
Db	1885	TTTTTGTATCTTAACAGTCACTAGTATTAATAATCTATTAAGC---GCTATATGTTTTTAAGA	1943
OY	1502	AAAAAAAAATCATTAACAGATCTATGAAGTAAGTAACATTTGATAGGTGCTCATTTAAAAA	1561
Db	1944	AAAAAAAAATCATTAACAGATTTATGAAGTAAGTAACATTTGATAGGTGCTCATTTAAAAA	2003

QY	1562	TAGTGGGATGATGCACAAATGCGTCTATGCTGTTGCGTGTACACTGAAATTAAC	162
Db	2004	TAGCTTGGAATGACAAATGCGCTCTATGCTGTTGCTGTGTACACTGAAATTAAC	206
QY	1622	AATATCTCTGATGATATAAAAAAAAAAAAAA	1653
Db	2064	AATATCTCTGATGATATAAAAAAAAAAAAAA	2095

		PRI	24-OCT-2000C
RESULT 7		mRNA	linear
AB050259			
LOCUS	AB050259	2140 bp	
DEFINITION	Macaca fascicularis brain cDNA,		
ACCESSION	AB050259		clone:QmPa-21045.
VERSION	AB050259.1 GI:10988236		
KEYWORDS	fat (full insert sequence).		
SOURCE	Macaca fascicularis adult male brain parietal lobe		cDNA to mRNA,
	clone_1lib:macaque brain cDNA library QmPa clone:		QmPa-21045.
ORGANISM	Macaca fascicularis		

REFERENCE AUTHORS	TITLE	JOURNAL REFERENCE AUTHORS	TITLE	JOURNAL REFERENCE AUTHORS
1 (sites)				
Osada, N., Hida, M., Kusuda, J., Tanuma, R., Iseki, K., Hirai, M., Terao, K., Suzuki, Y., Sugano, S. and Hashimoto, K.	Isolation of full-length cDNA clones from macaque brain cDNA libraries	2 (bases 1 to 2140)		
Hashimoto, K., Osada, N., Hida, M., Kusuda, J. and Sugano, S.		Direct Submission		
Submitted (20-Oct-2000)				

JOURNAL
Infectious Diseases, Division of Genetic Resources, 23-1, Toyama
1-chome, Shinjuku-ku, Tokyo 162-8640, Japan
(E-mail: khashi@nih.go.jp, URL: <http://www.nih.go.jp/yoken/genedbank/>
Tel: 81-3-5285-1111 (ex. 2120), Fax: 81-3-5285-1181)
Lab host: TOP10
COMMENT

FEATURES

Vectors: pME185-FU3 (Acc.No. AB009864)

R. Sites: DraIII (CAGCTGTGG).

R. Sites: DraIII (CACCATGTC)

Description: 1st strand cDNA was primed with an oligo(dT) primer [ATGCGCCCTTTTCTTTTTTTTTT]; double-stranded cDNA was synthesized using specific 5' and 3' primers and amplified by PCR. The PCR product was digested with SfiI and size selection was performed to exclude fragments <1.5kb. The SfiI-digested PCR product was cloned into distinct DraIII sites of pME185-FU3. XhoI sites just outside the DraII sites can be used to isolate the cDNA insert. Libraries were constructed by Sugano et al.(University of Tokyo, Institute of Medical Science). Custom primer used for sequencing
(5' end primer [CTTTCGCTTAACACTCGG];
3' end primer [CGACCTGCACCTCGACGACA]).

Location/Qualifiers

CDS

```

/organism="Macaca fascicularis"
/db_xref="taxon:9541"
/clone="Onpa-21045"
/sex="male"
/tissue_type="brain parietal lobe"
/clone_lib="macaque brain cdna library Onpa"
/dev_stage="adult"
110..1777
/codon_start=1
/product="MEPE protein"
/protein_id="BAB17010.1"
/db_xref="GI:10998327."
/translation="MRPECGLLLESTVAAPFQFQPTKTKSCVEQRITVYKGNH
KHHYVAVCYMSSGKQFQVQOEKKNNDNGLLHGLRRYQESKENVQRRKD
LSLEAEENNGSSQSYQFVFNRLRKEYSIKNGKINNGLRMSIYPSGNQFQAG
DLSLEAEENNGEAGALIRNMHOMINGPVTATIKLGEENKSKSKRYNTKIPSMYNA
KASHKDKKRFQDQSQFQPKQVYKSKSHRPHQNDIDPKLTSVKYKIPDFPESGCTDLD
ERGNDDKSPESGQFQPKQVQOEKKNNDNGLLHGLRRYQESKENVQRRKD
NEIPREKENGKNTIGQDEPAKADVAATSDLEGKQIQTFQFQPSAEASTLIDTKEPY

```

GONAHOCKVEFHYPPAPSKREKESDSESTNYNPIKNDKSAKAGVDSNRNOA
 ILHKOFPSKSGSLPIPSRGIDNITKEMSLNPSNENIPHSKXITVYHVRONM
 PTRNKMHPHKGSRQPSYNSRILSRSSRSDSSSSSSSDG*
 BASE COUNT 787 a 389 c 483 g 481 t
 ORIGIN

Query Match 91.1%; Score 1508.4; DB 9; Length 2140;
 Best Local Similarity 95.58; Pred. No. 0;
 Matches 1564; Conservative 0; Mismatches 71; Indels 3; Gaps 1;

2 TGAATAAGAAATATAGTATCAATTAACAAGAAATCTACATAGCGCTGAGATGTCAA 61
 489 TGAATAAGAAATATAGTATCAATTAACAAGAAATCTACATAGCGCTGAGATGTCAA 548
 62 TTTATCTTAAGTCAACTGGGAAATTAAGGTTTGAAGATGAGATGATCTATCAGCAAC 121
 549 TTTATCTTAAGTCAACTGGGAAATTAAGGTTTGAAGATGAGATGATCTATCAGCAAC 608
 122 TACATGACCAAGAAATATAGGCGGAGCTCTCATGAAATATACATGACATATATAGG 181
 609 TACATGACCAAGAAATATAGGCGGAGCTCTCATGAAATATACATGACATATATAGG 668
 182 GGCACGTACTGCTGATTAATCTCTGGGGGAAAGAAACAAAGAAACACACTAGGAATG 241
 669 GGCACGTACTGCTGATTAATCTCTGGGGGAAAGAAACAAAGAAACACTAGGAATG 728
 242 TTTTAAACATATATCCAGCAATATATATATATATATATATATATATATATATATAT 301
 729 TTTTAAACATATATCCAGCAATATATATATATATATATATATATATATATATATAT 788
 302 ACCCTCAAGAGATTTCCCAAGCCCAAGAAAGTCCAGTAAAGCAAAAGCAACCCATCTGA 361
 789 ACCCTCAAGAGATTTCCCAAGCCCAAGAAAGTCCAGTAAAGCAAAAGCAACCCATCTGA 848
 362 TTCAACACAAATGACTACCTAAACATCTCTCAAAAGTCAAAAGTCCCAAGTAT 421
 849 CTCACACAAATGACTACCTAAACATCTCTCAAAAGTCAAAAGTCCCAAGTAT 908
 422 TTGAAGGCGGGTTATACAGATCTTCAAGAGAGAGGAGCAATATATATCTCTTCA 481
 909 TTGAAGGCGGGTTATACAGATCTTCAAGAGAGAGGAGCAATATATATCTCTTCA 968
 482 GTGGGAGCGGCAACCTTTTAAGACATCTGCTAAAGAGAGTACTGCTCTGAC 541
 969 GTGGGAGCGGCAACCTTTTAAGACATCTGCTAAAGAGAGTACTGCTCTGAC 1028
 542 TAGAAGCAAGATATTTCAACAGGGTTTGCAGGCGCAAGTGAAGTACTGATC 601
 1029 TAGAAGCAAGATATTTCAACAGGGTTTGCAGGCGCAAGTGAAGTACTGATC 1088
 602 TTGACACAAAGAAAGCAGTTATATAGATATCCAGAGAGAGAAATGGTGAATA 661
 1089 TTGACACAAAGAAAGCAGTTATATAGATATCCAGAAAGAAAGAAATGGTGAATA 1148
 662 CCATTGGAATGAGTGAATCTGCGAAAGAGGAGCATCTGTTGATGTCAGCTTGTAG 721
 1149 CCATTGGAATGAGTGAATCTGCGAAAGAGGAGCATCTGTTGATGTCAGCTTGTAG 1208
 722 AGGGGCAACAGATATCATAGGAGTATCAATTTTAAGAGTCTCTCTGGAAGAGAA 781
 1209 AGGGGCAACAGATATCATAGGAGTATCAATTTTAAGAGTCTCTCTGGAAGAGAA 1268
 782 ACAGAGTGGATGCTGGCAGCAAAATGTCACCAAGGAGGTTGATTTTCAATACCTC 841
 1269 ACAGAGTGGATGCTGGCAGCAAAATGTCACCAAGGAGGTTGATTTTCAATACCTC 1328
 842 CTGACACCTCAAAAGAGAAAGAAAGGAGAGTATGATGACAGCTGAAAGTACCAC 901
 1329 CTGACACCTCAAAAGAGAAAGAAAGGAGAGTATGATGACAGCTGAAAGTACCAC 1388
 902 AT 961
 1389 AT 1448

QY 962 GGAACCAAGCAACCTTAATGAATAAACAAGGTTTCTAGTAAGGGCAAAAGTACGGCC 1021
 DB 1449 GGAACCAAGCAATTTTACATGAAAAAACAAGGTTTCTAGTAAGGGCAAAAGTACGGCC 1508
 QY 1022 TGCCCATTCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1081
 DB 1509 TGCCCATTCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1568
 QY 1082 GCGCCAGTCATGAAAT 1411
 DB 1569 GCGCCAGTCATGAAAT 1625
 QY 1142 AAAATATATTCACAGGAAATAGGTTATGCCAAGGAAAGGCTCTGCGGTAGACAAC 1201
 DB 1626 AAAATATATTCACAGGAAATAGGTTATGCCAAGGAAAGGCTCTGCGGTAGACAAC 1685
 QY 1202 CCCATTCACAGAGGTTTATGTTCCGTAGAGAGGATGACAGTATGATCATCTGACA 1261
 DB 1686 CATATTCACAGAGGCTTATGTTCCGTAGAGAGGATGACAGTATGATCATCTGACA 1745
 QY 1262 GTGGCAGTTCAAGTGAAGGAGGAGGAGTGTACTGTCCACAGAGAGTCTTACAGGAGT 1805
 DB 1746 GTGGCAGTTCAAGTGAAGGAGGAGGAGTGTACTGTCCACAGAGAGTCTTACAGGAGT 1865
 QY 1322 GTCTGAAGACCTGCTACCTGCTGATGATATAGAGAGACCACTGACAGTACAGC 1381
 DB 1806 GTCTGAAGACCTGCTACCTGCTGATGATATAGAGAGACCACTGACAGTACAGC 1865
 QY 1382 GTGAAGAGAGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1441
 DB 1866 GTGAAGAGAGAGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1925
 QY 1442 TTTTGTATCTTATATAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 1501
 DB 1926 TTTTGTATCTTATATAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 1985
 QY 1502 AAAAAAATATATACAGATCTATGAAATAGTAAATGATGATGATGATGATGATGATGAT 1561
 DB 1986 AAAAAAATATATACAGATCTATGAAATAGTAAATGATGATGATGATGATGATGATGAT 2045
 QY 1562 TAGTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1621
 DB 2046 TAGTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2105
 QY 1622 AATATCTCTGATGATTA 1639
 DB 2106 AATATCTCTGATGATTA 2123

RESULT 8
 AB060891 2259 bp mRNA linear PRI 13-JUN-2001
 LOCUS
 DEFINITION
 Macaca fascicularis brain cDNA clone:Qlra-13588, full insert
 sequence.
 ACCESSION
 AB060891
 VERSION
 AB060891.1 GI:13874559
 KEYWORDS
 Oligo capping; fls (full insert sequence).
 SOURCE
 Macaca fascicularis adult male temporal lobe right cDNA to mRNA,
 clone.lib:macaque brain cDNA library Qlra clone:Qlra-13588.
 ORGANISM
 Macaca fascicularis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 Cercopithecinae; Macaca.
 REFERENCE
 1 (sites)
 Osada, N., Hida, M., Kusuda, J., Tanuma, R., Iseki, K., Hirai, M.,
 Terao, K., Suzuki, Y., Sugano, S., and Hashimoto, K.
 Isolation of full-length cDNA clones from macaque brain cDNA
 libraries
 Unpublished
 JOURNAL
 2 (bases 1 to 2259)
 REFERENCE
 Hashimoto, K., Osada, N., Hida, M., Kusuda, J., and Sugano, S.
 TITLE
 Direct Submission

Db 1217 ACAGAGTGGATGTGGCGGCCAAACGCTCACCAGGGAAGGTGATTCATTAACCTTC 1276
 QY 842 CTGCACCCCTCAAAAGAGAAAGAAAGGAGAGTGTATGATGAGTGAAGTACCACT 901
 Db 1277 CTGCACCCCTCAAAAGAGAAAGAAAGGAGAGTGTATGATGAGTGAAGTACCACT 1336
 QY 902 ATAAATGAATTTCTTAAATGAGCAAGGAGTACAGAAAGGAGTGTATGATGAGTGA 961
 Db 1337 ATAAATGAATTTCTTAAATGAGCAAGGAGTGTATGAGTGAAGTGTATGAGTGA 1396
 QY 962 GGAACCAAGCAACTTAAATGAGAAAGGTTCTTGTATGAGGCAAAAGTCAAGGCC 1021
 Db 1397 GGAACCAAGCAACTTGTATGAGAAAGGTTCTTGTATGAGGCAAAAGTCAAGGCC 1456
 QY 1022 TCCCATTCCTCTCTGCTGTATGAGTGAATGCAAAAGCAAAAGTGTATGAGTGAAG 1081
 Db 1457 TCCCATTCCTCTCTGCTGTATGAGTGAATGCAAAAGCAAAAGTGTATGAGTGAAG 1516
 QY 1082 GCCCAGTGTAGATATATATACATGAGAGAAATATCATTTATGATGCCACAGAGAC 1141
 Db 1517 GCCCAGTGTAGATATATATACATGAGAGAAATATCATTTATGATGCCACAGAGAC 1573
 QY 1142 AAAATATATTTCTACAGGGAATTAAGGATGACCAAGGAGGCTCTGGGGTACACAC 1201
 Db 1574 AAAATATATTTCTACAGGGAATTAAGGATGACCAAGGAGGCTCTGGGGTACACAC 1633
 QY 1202 CCCATTCCAACAGGAGGTTAGTCCCTGTAAGAAAGGATGACAGTGTATGATGATGAT 1261
 Db 1634 CATTATTCACACAGAGGCTTGTATGAGTGAAGGAGGAGACAGTGTATGATGATGAT 1693
 QY 1262 GTGGCAGTTCAAGTGTAGAGAGCGATGTACTATCCACAGAGGTTCCAGCGGGGTGACA 1321
 Db 1694 GTGGCAGTTCAAGTGTAGAGAGCGATGTACTATCCACAGAGGTTCCAGCGGGGTGACA 1753
 QY 1322 GTCTGAAGCTCTGTACCTGTATGATGATGATGATGATGATGATGATGATGATGAT 1381
 Db 1754 GTCTGAAGCTCTGTACCTGTATGATGATGATGATGATGATGATGATGATGATGAT 1813
 QY 1382 GTGAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1441
 Db 1814 GTGAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1873
 QY 1442 TTTTTCCTATCTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1501
 Db 1874 TTTTTCCTATCTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1932
 QY 1502 AAAAAAATTCATTAAGATCTATGATGATGATGATGATGATGATGATGATGATGATGAT 1561
 Db 1933 AAAAAAATTCATTAAGATCTATGATGATGATGATGATGATGATGATGATGATGATGAT 1992
 QY 1562 TGTGTGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1621
 Db 1993 TGTGTGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2052
 QY 1622 AATATCTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1654
 Db 2053 AATATCTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2085

RESULT 10
 AF325916 1617 bp mRNA linear PRI 03-JUL-2001
 LOCUS AF325916
 DEFINITION Homo sapiens matrix extracellular phosphoglycoprotein precursor
 (MEPE) mRNA, complete cds.
 ACCESSION AF325916
 VERSION AF325916.1 GI:14586740
 KEYWORDS
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens.
 Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 1617)
 AUTHORS Argiro, L., Desbarats, M., Glorieux, F. H. and Ecarot, B.

TITLE Mepe, the gene encoding a tumor-secreted protein in oncogenic hypophosphatemic osteomalacia, is expressed in bone
 JOURNAL Genomics 74 (3), 342-351 (2001)
 MEDLINE 21309068
 PUBMED 11414762
 REFERENCE 2 (bases 1 to 1617)
 AUTHORS Argiro, L., Desbarats, M., Glorieux, F. H. and Ecarot, B.
 TITLE Direct Submission
 JOURNAL Submitted (04-DEC-2000) Genetics, Shriners Hospital, 1529 Cedar Avenue, Montreal, QC H3G 1A6, Canada
 FEATURES
 source
 1. 1617
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /cell_type="bone"
 1. 1617
 /gene="MEPE"
 40. 1617
 /gene="MEPE"
 /codon_start=1
 /product="matrix extracellular phosphoglycoprotein precursor"
 /protein_id="AAK70343.1"
 /db_xref="GI:14586741"
 /translation="MRYECVGLLFSYVMAAPTPQPTKRTKQSCVEBQREKKNKDNIGFHLGRINQELSKENITVOERKDLDSSEASNKSSKSNQYFTNRHLEKYSISNKEHTNGRLMSITPKSTGNKGFEDGDALISKIDHDEVECAALIRNMHOMIPVYIKLGEENKENTPRVNLILPASMNYAKAHKDKKKRQSDQAKSPKPSKSTRICHNIDYKHLISKYKRLPDSFGSYDLDQERGDNDSPFGDQPPKRDIDGKEATGPDLEKGDIDQTFAPSPSEASTHLDITPKRGWELPERENGNTIGTDEYAKEDADVSLVEGSDINDGSTNFKELPGRBNRVDAGSONAHQGVEPHPAPSKERKDGSSDAESTNYNEIPKNGKSTKRGKVDHSNRQATLANKRQRPSPKSGQLPIPSGLDNEIKNE MSFENGSPHENITTHGRKYHYVPHRQNNSTRNKMPQKSGMGRQPHSNRFRSRRDDSESDSGSSSESDGD"
 40. 87
 /gene="MEPE"
 88. 1614
 /gene="MEPE"
 /product="matrix extracellular phosphoglycoprotein"
 BASE COUNT 604 a 300 c 373 g 340 t
 ORIGIN
 Query Match 78.1%; Score 1392; DB 9; Length 1617;
 Best Local Similarity 100.0%; Pred. No. 1e-272;
 Matches 1292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TGAATTAAGATATAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 61
 Db 326 TGAATTAAGATATAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 385
 QY 62 TTTATCTTAAGTCACTGGGAATTAAGGTTGAGATGAGATGATGATGATGATGATGAT 121
 Db 386 TTTATCTTAAGTCACTGGGAATTAAGGTTGAGATGAGATGATGATGATGATGATGAT 445
 QY 122 TACATGACCAAGAGAAATGCGCGAGCTCTCATCGAATTAACATGCAACATTAATGG 181
 Db 446 TACATGACCAAGAGAAATGCGCGAGCTCTCATCGAATTAACATGCAACATTAATGG 505
 QY 182 GGCAGTGAAGTGTAACTCTCTGGGGAGAAAGAAAGAGAAACACACCTAGCAATG 241
 Db 506 GGCAGTGAAGTGTAACTCTCTGGGGAGAAAGAAAGAGAAACACACCTAGCAATG 565
 QY 242 TTCTTAACATTAATCCAGAGATGAATTTATGCTTAACACACCTGGAAGATTAAGA 301
 Db 566 TTCTTAACATTAATCCAGAGATGAATTTATGCTTAACACACCTGGAAGATTAAGA 625
 QY 302 AGCCTCAAGAGATTCACAGCCAGAGAAAGTCCAGATAAAGCAAAAGCACCATCGTA 361
 Db 626 AGCCTCAAGAGATTCACAGCCAGAGAAAGTCCAGATAAAGCAAAAGCACCATCGTA 685
 QY 362 TTCACACACATTTGACTACTTAACATCTCTCAAAAGTCAAAAGATCCCATGATTT 421
 Db 686 TTCACACACATTTGACTACTTAACATCTCTCAAAAGTCAAAAGATCCCATGATTT 745

RESULT 11					
AF530559					
LOCUS	AF530559	1651 bp	mRNA	linear	ROD 13-AUG-2002
DEFINITION	Rattus norvegicus osteoregulin-like protein mRNA, complete cds.				
ACCESSION	AF530559				
VERSION	AF530559.1	GI:22212815			
KEYWORDS					
SOURCE	Norway rat.				
ORGANISM	Rattus norvegicus				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;				

REFERENCE	Rattus.	1 (bases 1 to 1651)
AUTHORS	Wang, X., Hu, B. and Wang, Y.	
TITLE	Rattus norvegicus DNA sequence expressed in B4 cell line (possible subtype of osteoregulin)	
JOURNAL	Unpublished	
REFERENCE	2 (bases 1 to 1651)	
AUTHORS	Wang, X., Hu, B. and Wang, Y.	
TITLE	Direct Submission	
JOURNAL	Submitted (15-JUL-2002) Radiation Oncology, Thomas Jefferson University, 1020 Sansom St. Thompson Bldg. B-13, Philadelphia, PA 19107, USA	
FEATURES	Location/Qualifiers	
source	1..1651	
	/organism="Rattus norvegicus"	
	/strain="F344"	
	/db_xref="taxon:10116"	
	/cell_line="B4"	
CDS	14..1345	
	/codon_start=1	
	/product="osteoregulin-like protein"	
	/protein_id="AA04404.1"	
	/db_xref="GI:22212816"	
	/translation="MQLLSQNLKTEHEHEPRTGMKCHKLMEDSSSGNQNIIH SVKEPVEVGTGEGRDAPLHLDDNQDGLNLTITQIPKSLVLTGEVDSNKEK POSVLSVLPIDVHNTINDYSEDTEQNOQDITLLONSPOGSKTPPARSTHLYLHLP KILSDPEDSAPDLVANGDDVPEPSDGGHFMHTDPDGGVADSDPESSGHRVSS NVELVDPHNTGLSGNEIPGREGHIGVATRGTAOGAGSDVSDIVEGSEITGSKF RELDPEGNRVADSSONAHOGKVEHEFPDAPSKKRYGSEREHTGAGYVEIPKSSG GASKDAEESKGNVTLTFESORPFGKGGOSHSIIGNVKSFEEDSSNSLSREGTAIAHR RLSIPTRNRMGSDRSGMSMARSRPHPRHRYSTRORDSESSSSSSSSSD"	
BASE COUNT	529 a	384 c 433 g 305 t
ORIGIN		
Query Match	32.2%; Score 532.6; DB 10; Length 1651;	
Best Local Similarity	63.4%; Pred. No. 2.4e-106;	
Matches 1013; Conservative 0; Mismatches 489; Indels 97; Gaps 9;		
58	TCATTTATTCCTAAGTCACTGGGAAATPAAGGGTTTGAGATGAGATGATGCTATAGC	117
146	TCTGTGAAGCCTGAGCCCATGGTGGATAAAGAACAGAGGGTGGGAGATGCTCCCTT	205
118	AAACTACATGACCAAGAAGATATGGGCGACCTCATCAAAATACATCAACATATATA	177
206	CACCTGCTTGACCAAGACAGGAGGATGCCACCTCTCAGAAATATACCTACACCTGTA	265
178	ATGGGGCCAGTGAAGTAACTCCTGGGGGAGAGAAAACAAAGAGACACACCTAGG	237
266	AAGAGTGTGTGAGCGGGGACGAGTACAGAGCGACAGAAACAAAGAGAGAAACCTCAG	325
238	AATGTTTCAACATTAATCCCGACGCAAGTATGATTAATGCTAAGCACCTCGAAGATATA	297
326	AGTGTTCACCGTAAATTCACACAGATGTCCACATATCTAATACGACTCTCTGAAAGATACA	385
298	AAGAAGCCTCAACAGATATCCCAAGCCAGAAAAGTCCAGTAAATAACCAAGACCCAT	357
386	GAGAAACCAACAGAGGATCTACTACTOCAGAACAGCCAGGACAAACCAACACACCCCT	445
358	CGTATTCAACACAACTTACTACTTAAACATCTCTCTCAAAAGTCAAAAAAATCCCACT	417
446	CGGGCCCGACGACACAGCCTACCTTAACATCTCCGCCCAAAATCAGAAAAGATTCTAGT	505
418	GATTTTGAAGCAGCGGTTATAGATCTTCAACAGAGAGGGGACATGATATATCTCTCT	477
506	GACTTTCAGGACATGCTTCCCGACAGCTTCTAGTGAAGGGGGATTAATGATGTCCCTCT	565
478	TTCAAGTGGGACCGGCAACCTTTTAAAGACATTCCTGTGAAGAGAGAACTACTGTCCT	537
566	TTCAAGTGGAGATGAGCAAACTTTATGACACCTCCGACAGAGAGAGTGTCTGTTGATCT	625
538	GACCTAAGAGCAAGATATTTCAACAGAGGTTTTCAGAGGCCCAAGTGAAGCTGAGAGTACT	597
626	GATCCTCAAAACCTCAGCTGTGCACCCCTGTGT--CAGGCTCCACCAATGTGCGAGATTGTT	682

COMMENT

of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

Center: Baylor College of Medicine.

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: GMA

Center clone name: CH230-11B16

Summary Statistics

Sequencing vector: Plasmid

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 12061 bases at least Q40

Consensus quality: 126182 bases at least Q30

Consensus quality: 131508 bases at least Q20

NOTE: Estimated insert size may differ from sequence length

(see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).

NOTE: This is a 'working draft' sequence. It currently

consists of 62 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

1 1069: contig of 1069 bp in length
1070 1169: gap of unknown length
1170 2314: contig of 1145 bp in length
2315 2414: gap of unknown length
2415 3550: contig of 1136 bp in length
3551 4963: gap of unknown length
4964 5063: contig of 1133 bp in length
5064 6271: gap of unknown length
6272 6371: gap of unknown length
6372 7857: contig of 1486 bp in length
7858 9906: contig of 1049 bp in length
9907 9106: gap of unknown length
9107 10720: contig of 1614 bp in length
10721 10820: gap of unknown length
10821 11853: contig of 1033 bp in length
11854 11953: gap of unknown length
11954 13115: contig of 1162 bp in length
13116 13215: gap of unknown length
13216 14446: contig of 1131 bp in length
14447 15977: gap of unknown length
15978 16077: gap of unknown length
16079 17689: contig of 1612 bp in length
17690 18987: gap of unknown length
18988 19087: contig of 1198 bp in length
19089 21446: gap of unknown length
21447 21547: gap of unknown length
21548 23482: contig of 1936 bp in length
23483 25723: gap of unknown length
25724 25823: contig of 2141 bp in length
25824 27149: gap of unknown length
27150 28663: contig of 1326 bp in length
28664 28763: gap of unknown length
28764 30632: gap of unknown length
30633 32080: contig of 1869 bp in length
32081 32180: gap of unknown length
32181 34021: gap of unknown length
34022 34121: contig of 1841 bp in length
34122 36060: contig of 1939 bp in length

36061 36160: gap of unknown length
36161 37317: contig of 1157 bp in length
37318 37417: gap of unknown length
37418 39691: contig of 2274 bp in length
39692 39791: gap of unknown length
39792 41839: contig of 2048 bp in length
41840 41939: gap of unknown length
41940 43817: contig of 1878 bp in length
43818 43917: gap of unknown length
43918 46079: contig of 2162 bp in length
46080 46180: gap of unknown length
46181 48668: contig of 2489 bp in length
48669 50979: gap of unknown length
50980 51079: contig of 2211 bp in length
51080 53320: gap of unknown length
53321 53420: contig of 2241 bp in length
53421 55272: gap of unknown length
55273 55372: contig of 1852 bp in length
55373 57327: gap of unknown length
57328 57427: contig of 1955 bp in length
57428 59613: gap of unknown length
59614 59713: contig of 2186 bp in length
59714 61817: gap of unknown length
61818 61917: contig of 2103 bp in length
61918 64312: gap of unknown length
64313 66627: contig of 2396 bp in length
66628 66727: gap of unknown length
66729 69791: contig of 2215 bp in length
69792 69891: gap of unknown length
69892 74334: contig of 3064 bp in length
74335 74455: gap of unknown length
74456 77868: contig of 4463 bp in length
77869 77969: gap of unknown length
77970 80955: contig of 3414 bp in length
80956 81056: gap of unknown length
81057 84342: contig of 2987 bp in length
84343 84442: gap of unknown length
84443 86371: contig of 3287 bp in length
86372 86471: gap of unknown length
86472 87673: contig of 1929 bp in length
87674 87773: gap of unknown length
87774 91770: contig of 1202 bp in length
91771 91870: gap of unknown length
91871 94338: contig of 3997 bp in length
94339 94638: gap of unknown length
94639 97030: contig of 2668 bp in length
97031 97130: gap of unknown length
97131 103898: contig of 2392 bp in length
103899 103998: gap of unknown length
103999 107372: contig of 6768 bp in length
107373 107472: gap of unknown length
107473 110167: contig of 3374 bp in length
110168 110267: gap of unknown length
110269 114127: contig of 2695 bp in length
114128 114227: gap of unknown length
114229 118441: contig of 3860 bp in length
118442 118541: gap of unknown length
118542 121332: contig of 4214 bp in length
121333 121432: gap of unknown length
121433 127382: contig of 2791 bp in length
127383 127882: gap of unknown length
127884 132945: contig of 5850 bp in length
132946 133045: gap of unknown length
133046 139139: contig of 5563 bp in length
139140 139239: gap of unknown length
139240 142147: contig of 6094 bp in length
142148 142247: contig of 2908 bp in length

Query Match 31.5%; Score 522; DB 2; Length 169603;
Best Local Similarity 63.2%; Pred. No. 3.3e-104;
Matches 1000; Conservative 0; Mismatches 486; Indels 97; Gaps 9;

TITLE Direct Submission
JOURNAL Submitted (23-AUG-2000) Cardiovascular & Metabolic Disease, Pfizer
Global Research & Development, Eastern Point Road, Groton, CT
06340, USA

FEATURES Location/Qualifiers

source 1.1682
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="5"
/map="5E3-E5"
67, 1392
/note="extracellular matrix protein; expressed in osteoblasts and osteocytes of bone; similar to rat osteoblast/osteocyte factor 45"
/product="osteoregulin"
/protein_id="AA087687.1"
/db_xref="GI:21998851"
/translation="MPEGLMKQAVSVGLLFTSWTAAPMPNEDRSQNDISIKD
LAASVDPDPTVDEGTEGQALLHPGQDARYGALLNITQPVKSLVIGALELRREGNO
EKRPQSVLSVLPADVNDKSLDKINQESYLLTQSSPVSQKHKTRHQRSTHYLT
HLPOIKRTPSDLESGSPDLVRDNDVPEPFGDQHFHMI PGKGAAGSGPESSTSR
LSSSKRAEVIDPHMSGISNEIPGRREGHSAYATRDKAOGASAGSLVGSNEIT
GSTNRELPRKEGRINAGSONAHQGVFHHYPOVAAREVKGGEVHEHAGAGVETIK
SSKSSSKDDESKGNOLITLASRPFGRKKSQGPALPSHSLSREVSFNHYVTHQ
NNLPNGMSQSRWSPSRPNHRRASTORDESSSSGSSSESHGD"

CDS

BASE COUNT 510 a 401 c 432 g 339 t
ORIGIN

Query Match 29.3%; Score 485.2; DB 10; Length 1682;
Best Local Similarity 62.3%; Pred. No. 5.8e-96;
Matches 1017; Conservative 0; Mismatches 483; Indels 132; Gaps 11;

OY 16 AGTATCAGTACAAAGAGATACCTCAAGTGGCTGAGATGTCAATTATTCCTAAGTCA 75
DB 166 AGCTGGCGCATCAAGACAGATTCAAGAGCTGGCAGACTGTGTATCCTGATCCC 225
OY 76 ACTGGGAATTAAGGTTTGAGGATGAGATGCTATCAGCAACTACATGACCAAGA 135
DB 226 ACGGTGATGAGAGCAGACAGAGATGGGCAAGGTGCTCTTCAACCCGCTGGCAGAC 285
OY 136 GAATATGGCGAGCTCTCATCAGAAATATCAACATATATATGAGGCGCAGTGACTCG 195
DB 286 AGTATGTGTGCTGCTCTCTCTGAAATATCAAGCAGCCTGTAAAGTGTAGTGGG 345
OY 196 ATTAACTCTGGGGGAAAGAAAGAAAGAACACACCTAGGATGTCTTAACATATC 255
DB 346 GCCGAACAGGAGGAGGAAAGAACAGAGAGACCTCAGAGTGTCTTAAGCGCTAAT 405
OY 256 CCAGAGATGATTAATGCTAAAGCAGACCTGAGAGATTAAGAAAGGCTCAAGAGAT 315
DB 406 CCAGAGATGATTAATGCTAAAGTCTCTTAAGACATTAAGAAATCAAGAGATAT 465
OY 316 TCCCAAGCCAGAAAGTCCAGTAAAG-----CAAAAGCACCCTGATTTCAA 366
DB 466 CTGTAAACCCAGAGAGCCCGGTCAAAAGCAACACCAACACCCGCGAGACCGA 525
OY 367 CACAACTATGACTACCTTAAGATCTCTCAAAAGTCAAAAGATCCCAAGTATTTGA 426
DB 526 CGGACACACTACTACTGACATCTCCACAGATCAAGAAACAGTCCAGTACCTTGA 585
OY 427 GGCAGGGTATACAGATCTCAAGAGAGGAGCAATGATATATCTCTTCAAGTGG 486
DB 586 GGCAGTGGCTCCCGAGATTTCTAGTAGGAGAGATATGATGTCCCTCTTCAAGTGA 645
OY 487 GACGCGCAACTTTTAAGACATCTCTGTAAAGAGAGAGTACTAGTCTGACCTAGAA 546
DB 646 GATGGGCAACATTTATGACATCTCTGCAAGG---AGTGTGAGGCTGCTGCTGAA 702
OY 547 GCGAAAGATATTTCAACAGAGGTTTGACAGGCCAAGTGAAGTGAAGTACTATCTGAC 606
DB 703 AGCTCAAGTATGCGCC---CCTCTAGGCTCAGCAAAAGTGAAGTATTTAGACCCAT 759

OY 607 ACAAAAAGCCAGGTATTAATGAGATCCAGAGAGAGAGAAATGGTGGAAATACCAT 666
DB 760 ATGTGTAGACTAGGCTCTAATGAGATCCCGGGAGAGAAAGACATGGTGGCAGTCTAT 819
OY 667 GGAACATGAGGATGAACTGCGAAGAGAGCAGATCTCTGTATGTACCTTTGTAGAGGC 726
DB 820 GCAACGAGAGCAAAAGCTCCACAGGGGACGCTCTCAGAGTGGAGACCTGTGGGGGC 879
OY 727 AGCAAGATATCATGAGTATGACCAATTTTAAGAGCTCCCTGGAAAGAAAGAAACAGA 786
DB 880 AGCAATGAATACACAGGAGCAGCAATTTTCAAGAGCTCCCGGAAAGAAAGAAACAGA 939
OY 787 GTGATGTGGCAGCAAAATGCTCACAAAGAGAGTGTGATTCATTAACCTCTCTGCA 846
DB 940 ATTAATGCGGAGCCCAAAATGCTCATCAGAGAGAAATGATTTCTATCCCAAGTG 999
OY 847 CCTCAAAAGAGAAAGAGAGAGAGAGAGTATGATGAGCTGAAGTACCACTATAT 906
DB 1000 GCCTGAGAGAGAAAGGTAAAGGGGGGGGTGAGCATGCA---GGAGAGCTGTATCAAC 1056
OY 907 GAAATCTTAATAATGCGAAGAGCATGCCAAGAGGTGTATGATCATTTCAATAGAAC 966
DB 1057 GAAATCCCAAGAGCAGCAAAAGTATGCTGTAGCAAAAGTGCAGAAAGTCCAAAGGAGAC 1116
OY 967 CAAGAACCTTAATAATGAAAGAAAGAGTTCCTAGTAAAGGCAAAAGTCAAGGCTGCCC 1026
DB 1117 CAATTAACCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1176
OY 1027 ATTCCTCTGCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1086
DB 1177 CTGCCCTCTCAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1214
OY 1087 AGTATGAGATATTAATTAACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1146
DB 1215 -----AGAAACATTTATGTGTTTCCATGAGCAAAAT 1245
OY 1147 AATTCACAGCAATTAAGGATATGAGCAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1203
DB 1246 AATCTTAACCGAATTAAGAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 1305
OY 1204 CATTCACAGAGAGGTTTATGTTCCCGTGAAGAGATGACATGATGATGATGATGATGAT 1263
DB 1306 AATTCACAG 1362
OY 1264 GGCAGTTCAAGTGAAG 1323
DB 1363 GGGAGTTTATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1411
OY 1324 CTGAAGACCTGCTACCTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1383
DB 1412 -----CCCTGCTGTAGTCTGTGAG 1431
OY 1384 GAAAGAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1443
DB 1432 GAAAG 1489
OY 1444 TTTGCTATCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1503
DB 1490 TGTGCTATTTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1549
OY 1504 AAAAAATCATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1563
DB 1550 AAAAACTTCAATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1609
OY 1564 GTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1619
DB 1610 GTCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1669
OY 1620 ACAATATCTCTC 1631
DB 1670 ATATGCTGTGCTC 1681

Thu Apr 17 09:34:32 2003

Search completed: April 16, 2003, 19:03:30
job time : 5142 secs /

us-09-700-696b-1.rge

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 16, 2003, 07:46:14 ; Search time 405 Seconds
(without alignments)
9202.610 Million cell updates/sec

Title: US-09-700-696B-1

Perfect score: 1655
Sequence: 1 gtaataagaatacatagat.....ataaaaaaaaaaaaaa 1655

Scoring table: IDENTITY-NUC
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N_Geneseq.101002:*

1: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1655	100.0	1655	AA36447	CDNA encoding a po
2	1655	100.0	1655	AA36447	Truncated phosphat
3	1654	99.9	2013	AA36447	Phosphatoin polyp
4	1638	99.0	1876	AA36447	Human osteoregulin
5	1638	99.0	1969	AA36447	Human osteoregulin
6	1638	99.0	2019	AA36447	Human osteoregulin
7	1638	99.0	2112	AA36447	Human osteoregulin
8	1624.4	98.2	2031	AA36447	Human full-length
9	1369	82.7	1662	ABA99160	Human phosphatoin

10	1289	77.9	1575	24	ABA99159	Human phosphatoin
11	514.8	31.1	1655	22	AA36447	Rat osteoregulin c
12	485.2	29.3	1682	22	AA36447	Mouse osteoregulin
13	425	25.7	501	22	AA36447	Human CDNA 3'-end
14	326.4	19.7	807	22	AA36447	Human CDNA 5'-end
15	326.4	19.7	807	22	AA36447	Human CDNA clone r
16	117.4	7.1	146	16	ABN23634	Human gene signatu
17	60	3.6	60	24	ABN23634	Human spliced tran
18	59.2	3.6	4590	22	AA36447	Yeast ADD9604-asso
19	50.8	3.1	5586	24	ABK40004	Human chemically p
20	47.6	2.9	327	23	ABV08167	Human prostate exp
21	47.4	2.9	2277	19	AAV05370	DNA encoding novel
22	47.2	2.9	1083	23	AA36447	Human chemically m
23	46.8	2.8	2277	19	AAV13834	Human chemically m
24	46.8	2.8	6699	24	ABN80183	Tumour suppressor
25	46.8	2.8	34769	22	AA36447	Human polynucleoti
26	46.6	2.8	414	22	AA188845	Human angiogenesis
27	46.6	2.8	7041	24	ABN80067	Human chemically m
28	46.4	2.8	6676	24	ABO67013	Plasmodium falcipa
29	45.6	2.8	5145	21	AA36447	Gene encoding a su
30	45.6	2.8	6741	21	AA36447	Tomato spotted wil
31	45.4	2.7	1710	15	AA36447	Chemically treated
32	45.4	2.7	2028	15	AA36447	Human transductio
33	44.8	2.7	12870	24	ABR39983	Tumour suppressor
34	44.8	2.7	12870	24	ABR39983	Human immune syste
35	44.8	2.7	47108	24	ABR31511	Human immune syste
36	44.6	2.7	14615	22	AA36447	Human immune syste
37	44.4	2.7	5218	24	ABL33266	Human immune syste
38	44.4	2.7	14006	24	ABL33266	Human immune syste
39	44.2	2.7	8530	24	ABR31414	Signal transductio
40	44.2	2.7	8530	24	ABR31414	Human immune syste
41	44.2	2.7	21537	24	ABL33998	Human angiolegenis
42	44.2	2.7	5237	24	ABO67107	DNA transcription
43	44.4	2.7	6155	24	ABR28361	Human angiolegenis
44	44.4	2.7	9095	24	ABO67061	Human immune syste
45	44.4	2.7	9997	24	ABL32369	Human immune syste

ALIGNMENTS

RESULT 1

AA36447 standard; CDNA: 1655 BP.

AA36447:

22-FEB-2000 (first entry)

CDNA encoding a polypeptide designated phosphatoin.

Human: phosphatoin; Metastatic-tumour Excreted Phosphatidic-Acetyl-CoA;
MEPE: Na+-dependent phosphate cotransport; vitamin D metabolism;
bone mineralisation; phosphate metabolism related disease;
hyperphosphatemia; renal osteodystrophy; renal dialysis;
secondary hyperparathyroidism; osteitis fibrosa cystica; hypercalcaemia;
X-linked hypophosphatemic rickets; hereditary hypophosphatemic rickets;
hypomineralised bone lesion; stunted growth; cystic fibrosis;
oncogenic hypophosphatemic osteomalacia; renal phosphate leakage;
renal osteodystrophy; osteoporosis; vitamin D resistant rickets;
renal organ resistance; renal Fanconi syndrome; autosomal rickets;
Paget's disease; kidney failure; renal tubular acidosis; spinae; ds.

Homo sapiens.

Key Location/Qualifiers

FT 1..1293

FT CDS

FT /tag= a

FT /product= "phosphatoin"

FT /note= "5' end of the sequence is not given"

FT polyA_signal

FT 1613..1620

FT /tag= b

XX

PN WO9960017-A2.

PD 25-NOV-1999.

PF 18-MAY-1999; 99WO-EP03403.
YY

PR	18-MAY-1998;	98GB-0010681.
PR	04-SEP-1999	98GB-0010337

PK 04-SEP-1998; 98GB-0019387.
YY

PA (UNLO) UNIV COLLEGE LONDON.
XY

Rowe P;

DR WPI; 2000-053262/04.

P-PSDB; AAY53812.

PT New polypeptides involved in the regulation of phosphate metabolism
PT useful for diagnosing and treating disorders related to phosphate
PT metabolism -

PS Claim 6; Fig 8; 136pp; English.
XY

Metastatic sequence encodes a phosphatonin polypeptide (also called phosphatonin in a subject Excreted Phosphaturic Element (MPE)). The level of vitamin D metabolism and/or bone mineralisation. The phosphatonin polypeptides, polynucleotides, vectors and antibodies are used to treat phosphate metabolism related disease. They are used for treatment of dialysis/pre-dialysis, secondary hyperparathyroidism or osteitis fibrosa cystica, or X-linked hypophosphatemic rickets, hereditary hypophosphatemic rickets with hypocalcaemia (HHR), hypomineralised bone lesions, stunted growth in juveniles, oncogenic hypophosphatemic osteomalacia, renal phosphate leakage, renal osteodystrophy, Fanconi syndrome, autosomal rickets, end organ resistance, renal tubular acidosis, cystic fibrosis or sprue. The polypeptide may also be used to manufacture combined preparations for simultaneous, separate or sequential use for the treatment of phosphate metabolism disorders. A transformed osteoblast or bone cell line capable of phosphatonin overexpression is useful for the production of phosphatonin.

Sequence 1655 BP; 609 A; 303 C; 380 G; 363 T; 0 other;

Query Match	100.0%;	Score 1655;	DB 21;	Length 1655;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1655;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

1 GTCAATTAAGAAATATAGTATCAGTAAACAAAGAAATCTCAATATGGCTGAGATGTCCA 60
|||||
1 GTCAATTAAGAAATATAGTATCAGTAAACAAAGAAATCTCAATATGGCTGAGATGTCCA 60

61 ATTATTCCTAAGTCAACTGGGAATAAAGCGTTTGAGGATGGAGATGATGCTTATCAGCAA 120

121 CTACATGACCAAGAGAATATGCGCGACGCTCTCATCAGAAATACATGCAACATATATG 180
|||||
121 CTACATGACCAAGAGAATATGCGCGACGCTCTCATCAGAAATACATGCAACATATATG 180

181 GGGCCAGTGACTGGCATTAACTCCTGGGGGAAGAAACAAAGAGAACACACCTTAGAAT 240
|||||
181 GGGCCAGTGACTGGCATTAACTCCTGGGGGAAGAAACAAAGAGAACACACCTTAGAAT 240
GGGCCAGTGACTGGCATTAACTCCTGGGGGAAGAAACAAAGAGAACACACCTTAGAAT 240

241 GTTCTAAACATATATCCAGCAAGTATGAAATTATGCTAAAGCACACCTGGAAGCATAAAAG 300
|||||
241 GTTCTAAACATATATCCAGCAAGTATGAAATTATGCTAAAGCACACCTGGAAGCATAAAAG 300

301 AAGCTCAAGAGATTECCAAAGCCAGAAAAGTCACATAAAAGCAAAAGCACCACCTCGT 360
|||||
301 AAGCTCAAGAGATTECCAAAGCCAGAAAAGTCACATAAAAGCAAAAGCACCACCTCGT 360
|||||

```
QY      361 ATTCAACGACACATTTGACTACCTTAAACCATCTGTCAAAGTGCAAAAAAATCCCGAGTAT   420
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      361 ATTCACGACACATTGACTACCTTAACCATCTGTCAAAGTGCAAAAAAATCCCGAGTAT   420
```

421	TTTGAAGCAGCGGTTATACAGATCTTCAAGAGAGAGGGACATGATATATCTCTTC	480
421	TTTTGAAGCAGCGGTTATACAGATCTTCAAGAGAGAGGGACATGATATATCTCTTC	480

QY 481 AGTGGGACGGCCAACTTTTAAAGCATTCTGTGTTAAGGAGAGCTACTGTCCTGAC 540
|||||
Dδ 481 AGTGGGACGGCCAACTTTTAAAGCATTCTGTGTTAAGGAGAGCTACTGTCCTGAC 540
|||||

QY	541 CTGAAGCGCAAAAGATATTTCACAACAGGGTTTGCAGGCCCAAGTGAAAGCTGAGACTACTCAT 600
Db	541 CTGAAGCGCAAAAGATATTTCACAACAGGGTTTGCAGGCCCAAGTGAAAGCTGAGACTACTCAT 600

QY 601 CTTGACACAAAAAGCCAGGTTATATATGAGATCCAGAGAGAGAGAAGAAATGCTGGAAT 660
|||||
601 CTTGACACAAAAAGCCAGGTTATATATGAGATCCAGAGAGAGAGAAGAAATGCTGGAAT 660
|||||

661 ACCATTGGAACTAGGATGAACCTCGGAAGAAGGCGCATGCTTTGATGTACAGCTTCTA 720

721 GAGGCGCAGCAAGATATCATGCGTAGTACCAATTTTAAGAGACTCCTGGAGAGAGAGCA 780

781 AACAGAGTGGATGCTGGCAGCCAAATGCTCACCAAGGGAAGTTTACATTACCT 840
|||||

841 CCTGCACCCCTCAAAAGAGGAAAAAGAGGAGGCTAGTCATGCTGAAGTACCAAC 900

841 CCTGACCCCTCAAAAAGAGAAAAAGGAGGAGTGTGATGCAGCTGAAGTACCAAC 900

901 TATATGAAATTCCTAAAAATGGCAAGCAGTACAGAAAGGCTGTAGATCATTTCTAAT 960

901 TATATGAATTCCTAAAAATGGCAAGGACGATCCAGAAAGGGTGTAGATCATTTCTAAT 960
961 AGGAACCAAGCAACCTTAAATGAAAAACAAGGTTTCCCTAGTAAGGGCAAAAGTCAGGGC 1020

1021 CTGCCCATTTCTTCTTCGCTGCTTGAATATGAAATCAAAATCCAAATTCCTATCCCTCTTTT 1000
961 AGGAACCAAGCAACCTTAAATGAAACCAAGGTTTCTAGTAAGGCGAAAGTCAAGGC 1020

b

1081 GGGCCCACTGATCAGATAATTATTTACCTTAA
1021 CTGCCCATTCCTTCGTGCTTGATTAAGAATCAA
1081

1081 GCGCCGTCATGAGATTAATACACATGCGAGAAATATCTTGTATCCCGACAGA 1140

1141 CAAATATTTCTACACGGAATAGGGTATGCCACAGGGAAGGCTCTGGGGTAGACAA 1200

1201 CCCCATCCCAACAGGAGGTTAGTTCCTCCCGTAGAAGGGATGACAGTAGAGTCACTCTGAC 1260
|||||
1201 CCCCATCCCAACAGGAGGTTAGTTCCTCCCGTAGAAGGGATGACAGTAGAGTCACTCTGAC 1260

1261 AGTGCAGTTCAAAGTGAGAGCGATGGTGAAGTGTAGTCCACAGAGATTCCAGCGGGGTGAC 1320

1261 AGTGCAGTTCAAAGTGAGAGCGATGGTGAAGTGTAGTCCACAGAGATTCCAGCGGGGTGAC 1320

1321 AGTCTGAAGACCTCGTCACTCTGTGATTATGTAGAGAGAGCCACTGACACTGACCA 1380
|||||
1321 AGTCTGAAGACCTCGTCACTCTGTGATTATGTAGAGAGAGCCACTGACACTGACCA 1380
|||||

1381 GGTCAGAGAGGATAGAGTGAAGAACTAGTGAGCCAAAGATCTGGTCTCTGGGGGA 1440
1441 ATTTTTCCTACTTAATAGTCACAGTATAAAATTCCTATTAAAGGCTATAATGTTTTTAAG 1500


```

OY 901 TATATGAATTCCTAAAAATGGCAAGGCACTACGAGAAAGGTGTAGATCATTTCTAT 960
    |||||||
DB 901 TATATGAATTCCTAAAAATGGCAAGGCACTACGAGAAAGGTGTAGATCATTTCTAT 960
OY 961 AGGAACCAAGCACTTAAATGAAAAACAAGCTTCTAGTAAGGGCAAAAGTCAGGGC 1020
    |||||||
DB 961 AGGAACCAAGCACTTAAATGAAAAACAAGCTTCTAGTAAGGGCAAAAGTCAGGGC 1020
OY 1021 CTGCCATTCCTTCGTGCTGTGTATGTAATCAAAATGAAATGGATTCCTTAT 1080
    |||||||
DB 1021 CTGCCATTCCTTCGTGCTGTGTATGTAATCAAAATGAAATGGATTCCTTAT 1080
OY 1081 GGGCCCACTCATGAGAAATATAACATGAGCAAAATATCATTAATGTACCCACAGA 1140
    |||||||
DB 1081 GGGCCCACTCATGAGAAATATAACATGAGCAAAATATCATTAATGTACCCACAGA 1140
OY 1141 CAAATTAATTTCTACACGGAATTAAGGATATCCACAAGGAAAGGCTCTGGGGTAGACAA 1200
    |||||||
DB 1141 CAAATTAATTTCTACACGGAATTAAGGATATCCACAAGGAAAGGCTCTGGGGTAGACAA 1200
OY 1201 CCCCATCCCAACAGAGGTTTATGTTCCCGTAGAAGGAGTACAGTAGTACATCTGAC 1260
    |||||||
DB 1201 CCCCATCCCAACAGAGGTTTATGTTCCCGTAGAAGGAGTACAGTAGTACATCTGAC 1260
OY 1261 AGTGCACTTCAAGTAGAGAGCGATGTGACTAGTCCACAGAGTCCACAGCGGGTGAC 1320
    |||||||
DB 1261 AGTGCACTTCAAGTAGAGAGCGATGTGACTAGTCCACAGAGTCCACAGCGGGTGAC 1320
OY 1321 AGTCTGAAGACCTCTCTACACCTGTGTAGTGTAGAGAGAGCCACTGACACTGACCA 1380
    |||||||
DB 1321 AGTCTGAAGACCTCTCTACACCTGTGTAGTGTAGAGAGAGCCACTGACACTGACCA 1380
OY 1381 GGTGAGAGAGAGATAGTAGTAAGAACTGATGAGGCAAGAAATCCCTGGTCTCTGGGGGA 1440
    |||||||
DB 1381 GGTGAGAGAGAGATAGTAGTAAGAACTGATGAGGCAAGAAATCCCTGGTCTCTGGGGGA 1440
OY 1441 ATTTTGTCTATCTTAATAGTACAGATATAAAATCTATTAAGGCTATTAATGTTTTTAA 1500
    |||||||
DB 1441 ATTTTGTCTATCTTAATAGTACAGATATAAAATCTATTAAGGCTATTAATGTTTTTAA 1500
OY 1501 CAAAAAAAATCATTTACAGATCTATGAATAGTAACTTTGAGTAGTGTCAATTTAA 1560
    |||||||
DB 1501 CAAAAAAAATCATTTACAGATCTATGAATAGTAACTTTGAGTAGTGTCAATTTAA 1560
OY 1561 ATAGTTGTGATGTACACAATGCGCTTATGTTGTTGCTCTGTAGACATGAAAAATAAA 1620
    |||||||
DB 1561 ATAGTTGTGATGTACACAATGCGCTTATGTTGTTGCTCTGTAGACATGAAAAATAAA 1620
OY 1621 CAATATCTCTCGATGATTAATAAAAAAAAAAAAAAAAAAAAA 1655
    |||||||
DB 1621 CAATATCTCTCGATGATTAATAAAAAAAAAAAAAAAAAAAAA 1655

```

RESULT 3
AAF83764
ID AAF83764 standard; cDNA: 2013 BP.
AAF83764;

DT 23-JUL-2001 (first entry)
XX Phosphatonin polypeptide (MEPE) encoding cDNA.
XX Metastatic-tumour excreted phosphaturic element; MEPE: phosphatonin;
KM phosphate; vitamin-D; skeletal formation; mineralization; truncated;
KM osteopathic; antihypertensive; cytoskeletal; human; ss.
XX Homo sapiens.
XX Key Location/Qualifiers
FH CDS 74..1651
FT /+tag- a

```

FT /product= "phosphatonin"
XX MO200112878-42.
PN 10-MAY-2001.
XX 31-OCT-2000; 2000WO-EPI0747.
XX 04-NOV-1999; 99US-0434185.
PR 08-NOV-1999; 99GB-0026424.
XX (UNIO ) UNIV COLLEGE LONDON.
XX Rowe P.
XX WPI: 2001-343487/36.
DR P-PSDB; AAB62689.
XX
XX New phosphatonin polypeptide a regulator of phosphate metabolism, for
XX diagnosing and treating disorders of phosphate, vitamin-D metabolism,
XX skeletal formation e.g. osteoporosis, Paget's disease, gout
XX
XX Examples; Page 132-133; 135pp; English.
XX
XX The invention relates to a novel human protein, metastatic-tumour
XX excreted phosphaturic element (MEPE) or phosphatonin (modulator of
XX phosphate and vitamin-D metabolism). The phosphatonin polypeptides,
XX polynucleotides and specific antibodies are useful for treating a
XX disorder of phosphate or vitamin D metabolism, skeletal formation and
XX mineralization. Phosphatonins are used to treat hyperphosphatemia, renal
XX osteodystrophy, secondary hyperparathyroidism, osteitis fibrosa cystica
XX or gout. It is used to prepare a medicament for treating X-linked
XX hypophosphatemic rickets, hereditary hypophosphatemic rickets with
XX hypercalcaemia (HHR), hypomineralized bone lesions, stunted growth in
XX juveniles, oncogenic hypophosphatemic osteomalacia, renal phosphate
XX leakage, renal osteodystrophy, osteoporosis, vitamin-D resistant rickets,
XX end organ resistance, renal Fanconi syndrome, autosomal rickets, Paget's
XX disease, kidney failure, renal tubular acidosis, cystic fibrosis or
XX sprue. Phosphatonin polynucleotides are useful as molecular weight
XX markers on Southern gels, as diagnostic probes for detecting the presence
XX of a specific mRNA. Phosphatonin polypeptides are also useful for
XX identifying agonists and antagonists, compounds which bind to
XX phosphatonin and drug candidates for therapy of phosphate metabolism
XX disorders. The present sequence represents the nucleotide sequence of a
XX second cloned cDNA for the entire phosphatonin (MEPE).
XX
XX Sequence 2013 BP; 740 A; 367 C; 459 G; 447 T; 0 other;
XX
XX Query Match 99.9%; Score 1654; DB 22; Length 2013;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 1654; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
OY 2 TGAATTAAGATATAGTATCAGTAACAAAGAGATCTCAATGCGCTGAGATGTCAA 61
    |||||||
DB 360 TGAATTAAGATATAGTATCAGTAACAAAGAGATCTCAATGCGCTGAGATGTCAA 419
OY 62 TTTATCTAGTCACTGGGGAATTAAGGTTTGAGTAGAGATGATGATGAGCAAC 121
    |||||||
DB 420 TTTATCTAGTCACTGGGGAATTAAGGTTTGAGTAGAGATGATGATGAGCAAC 479
OY 122 TACATGACCAAGAAAGATATGCGCAGCTCTCATCAGAAATACATGCAATATATAG 181
    |||||||
DB 480 TACATGACCAAGAAAGATATGCGCAGCTCTCATCAGAAATACATGCAATATATAG 539
OY 182 GGCAGAGTCTGGATTAACCTCTGGGGGAAGAAACAAAGAGACACCTAGAGATG 241
    |||||||
DB 540 GGCAGAGTCTGGATTAACCTCTGGGGGAAGAAACAAAGAGACACCTAGAGATG 599
OY 242 TTCTAAACATATCCGACAGATATGATATGCTAAGACACTGGAAGATTAAGA 301
    |||||||
DB 600 TTCTAAACATATCCGACAGATATGATATGCTAAGACACTGGAAGATTAAGA 659
OY 302 AGCCTCAAGAGATTCCTCAAGCCGAGAAAGTCCAGTAAAGCAAAAGCAACCATCGTA 361

```


OY	2	TGAATTAAGAAATATATGTATACCTAATCAAAAGAAATCTCACATAGGCTGAGAGATTCTCA	61
Db	239	TGCAATTAAGAAATATATGTATACCTAATCAAAAGAAATATCTACATATGGCTGAGAGATTCTCA	298
OY	62	TTTATCTTAAGTCAACTGGGAAATTAAGGGTTTGAGAGATGAGATGCTATACGCAAC	121
Db	299	TTTATCTTAAGTCAACTGGGAAATTAAGGGTTTGAGAGATGAGATGCTATACGCAAC	358
OY	122	TACATGACCAATAAGAAATTTGGCGAGCTCTCATCGAATTAACATATGCAACATATATATG	181
Db	359	TACATGACCAATAAGAAATTTGGCGAGCTCTCATCGAATTAACATATGCAACATATATATG	418
OY	182	GGCCAGTACTGCGATTAAACTCTGGGGGGAAGAAAACAAGAGAACACACTAGGAATG	241
Db	419	GGCCAGTACTGCGATTAAACTCTGGGGGGAAGAAAACAAGAGAACACACTAGGAATG	478
OY	242	TTCTAATCACTATCCAGCAAGTATGAATTAATGCTTAACACACTGGAAGATTAATAAG	301
Db	479	TTCTAATCACTATCCAGCAAGTATGAATTAATGCTTAACACACTGGAAGATTAATAAG	538
OY	302	AGCCTCAAAAGAGATTCCCAAGCCAGAAAAAGTCCAGTAAAAAAGCAAAAGCACCCATCGTA	361
Db	539	AGCCTCAAAAGAGATTCCCAAGCCAGAAAAAGTCCAGTAAAAAAGCAAAAGCACCCATCGTA	598
OY	362	TTTCAACACACATTCACATCTTAAACATCTCTCAAAAGTCAAAAAAATCCCGAGTAT	421
Db	599	TTTCAACACACATTCACATCTTAAACATCTCTCAAAAGTCAAAAAAATCCCGAGTAT	658
OY	422	TTGAAGGCGAGCGTTATACAGATCTTCAAGAGAGAGGGGACATATATCTCTTTCA	481
Db	659	TTGAAGGCGAGCGTTATACAGATCTTCAAGAGAGAGGGGACATATATCTCTTTCA	718
OY	482	GTGGGAGCGGCCAACCTTTTAAGACATCTCTGTAAAGAGAGAGTACTGTCCTGACC	541
Db	719	GTGGGAGCGGCCAACCTTTTAAGAGATCTCTGTAAAGAGAGAGTACTGTCCTGACC	778
OY	542	TGAAGGCGAAAGATATTTCAACAGGGTTTGCAGGGCCCAAGTGAAGCTGAGACTCATC	601
Db	779	TGAAGGCGAAAGATATTTCAACAGGGTTTGCAGGGCCCAAGTGAAGCTGAGACTCATC	838
OY	602	TTGACACCAAAAAAGCGAGTTATTAATGAGATCCACAGAGAGAGAAATGGTGGAAAT	661
Db	839	TTGACACCAAAAAAGCGAGTTATTAATGAGATCCACAGAGAGAGAAATGGTGGAAAT	898
OY	662	CCATTGGAAGTATGAGATGTAACCTGGCAAAAGAGGCAATCTGTTGATGTCAGCCCTGTAG	721
Db	899	CCATTGGAAGTATGAGATGTAACCTGGCAAAAGAGGCAATCTGTTGATGTCAGCCCTGTAG	958
OY	722	AGGCGACGACAGATATCATGCGTAGTACCAATTTTAAGAGATCTCCCTGGAAAGAAAGAA	781
Db	959	AGGCGACGACAGATATCATGCGTAGTACCAATTTTAAGAGATCTCCCTGGAAAGAAAGAA	1018
OY	782	ACAGAGTGAATGCTGGCAGCCAAATATGCTCACCAAGGAAAGGTTGAGTTTCATTACCTC	841
Db	1019	ACAGAGTGAATGCTGGCAGCCAAATATGCTCACCAAGGAAAGGTTGAGTTTCATTACCTC	1078
OY	842	CTTGCAACCTCAAAAGAGAAAAGAAAGACAGTAGTATGAGCTGAAGTACCAACT	901

Db	1079	CTGACCCCTCAAAAAGSAAAAGAAAAGGAAGGACGATGATGATGCACCTGAAAGTACCAACT	1138
Qy	902	ATAATGAATTCCTAAAAATGSCCAAGGACGATPCCAGAAAGGCTGATCATATTCCTATA	961
Db	1139	ATAATGAATTCCTAAAAATGSCCAAGGACGATPCCAGAAAGGCTGATCATATTCCTATA	1198
Qy	962	GGAACCAAGCAACCTTAAATGAAGAAAACAAAGGTTCTGATGAAGGCAAAAGTCAGGGCC	1021
Db	1199	GGAACCAAGCAACCTTAAATGAAGAAAACAAAGGTTCTGATGAAGGCAAAAGTCAGGGCC	1258
Qy	1022	TGCCATTCCTCCGTCGTCGTGATGAATGAATCAAAAACGAATGCAATTCCTTAAATG	1081
Db	1259	TGCCATTCCTCTCTCGGTGCTTGATATGATGAATCAAAAACGAATGCAATTCCTTAAATG	1318
Qy	1082	GCCCAGTCATGAGATATATATATACACATGCGCAGAAATATCATTTATGATACCCACAGAC	1141
Db	1319	GCCCAGTCATGAGATATATATATATACACATGCGCAGAAATATCATTTATGATACCCACAGAC	1378
Qy	1142	AAATATATTTCTACACGGSAATGAAGGATATGCCAAGGGGAAAGGCTCCGCGGGGTGACACAC	1201
Db	1379	AAATATATTTCTACACGGSAATGAAGGATATGCCAAGGGGAAAGGCTCCGCGGGGTGACACAC	1438
Qy	1202	CCCATTCCTCAACAGAGAGGTGTAGTTCCCGTAGAAGAGGATACAGTATGATCATCTGACA	1261
Db	1439	CCCATTCCTCAACAGAGAGGTGTAGTTCCCGTAGAAGAGGATACAGTATGATCATCTGACA	1498
Qy	1262	GTGCGACTTCAAGTGAAGACGGATGTTGACTAGTCCACACAGAGTTCCACAGGGGGGTGACA	1321
Db	1499	GTGCGACTTCAAGTGAAGACGGATGTTGACTAGTCCACACAGAGTTCCACAGGGGGGTGACA	1558
Qy	1322	GTCGGAAGACCTCGTCACTCGTATAGTTGATGTGAGAGAGAGGCACTCACAAGCTGACCAG	1381
Db	1559	GTCGGAAGACCTCGTCACTCGTATAGTTGATGTGAGAGAGAGGCACTCACAAGCTGACCAG	1618
Qy	1382	GTGAAGAGAGATAGACTGTAAGAAGTGAAGTGAAGCCACAAGATCCTGCTCTCTTGGGGGAA	1441
Db	1619	GTGAAGAGAGATAGACTGTAAGAAGTGAAGTGAAGCCACAAGATCCTGCTCTCTTGGGGGAA	1678
Qy	1442	TTTTTGTGCTATCTTAATAGTCAACAGTATAAAATTTCTATTAAAGCTATATATGTTTTAAGC	1501
Db	1679	TTTTTGTGCTATCTTAATAGTCAACAGTATAAAATTTCTATTAAAGCTATATATGTTTTAAGC	1738
Qy	1502	AAAAAAAATCATATACAGATCTATGAAATATAGTAACATTGAGTATAGTATGATTTAAATA	1561
Db	1739	AAAAAAAATCATATACAGATCTATGAAATATAGTAACATTGAGTATAGTATGATTTAAATA	1798
Qy	1562	TAGTTGGTGAATGTCACAAAATGCTCTATATGTTGTTGCTCTGTAGACATGAAATTAAC	1621
Db	1799	TAGTTGGTGAATGTCACAAAATGCTCTATATGTTGTTGCTCTGTAGACATGAAATTAAC	1858
Qy	1622	AATATCTCTCGATGATGA	1639
Db	1859	AATATCTCTCGATGATGA	1876

	RESULT 5
AAH26811	
ID	AAH26811 standard; cDNA; 1969 BP.
XX	
AC	AAH26811;
XX	
DT	21-DEC-2001 (first entry)
XX	
DE	Human osteoregulin (mature polypeptide) cDNA.
XX	
KW	Osteoregulin; human; bone; homeostasis; adipose; calcification;
KW	atherosclerosis; osteoporosis; osteopathic; antiarteriosclerotic;
KW	therapy; ss.
XX	
OS	Homo sapiens.
XX	
XX	
FT	Key Location/Qualifiers
	mat_peptide 1..1966

```

FT      /*tag- a
XX      EP130098-A2.
PN      05-SEP-2001.
XX      27-FEB-2001; 2001EP-0301768.
XX      29-FEB-2000; 2000US-185617P.
PR      22-SEP-2000; 2000US-234500P.
XX      (PF12 ) PFIZER PROD INC.
XX      Brown TA, De Wet JR, Gowen LC, Hames LM;
XX      WPI: 2001-604111/69.
DR      P-PSDB: AAB82923.
XX      Novel osteoregulin polypeptide useful for regulating bone homeostasis,
PT      adiposity and calcification of atherosclerotic plaques comprises
PT      measuring the activity of osteoregulin -
XX      Claim 4; Page 58-59; 90pp: English.
PS      The present sequence is that of cDNA encoding human osteoregulin
CC      mature polypeptide (see AAB82923). It is derived from an
CC      osteoblast cDNA clone (see AAB826809), which encodes an osteoregulin
CC      polypeptide (see AAB82921) including an N-terminal signal peptide.
CC      Osteoregulin is a novel protein which plays a role in regulating bone
CC      homeostasis, adiposity, and the calcification of atherosclerotic
CC      plaques. A splice variant of this sequence (see AAB82922) was also
CC      identified. The invention provides osteoregulin proteins, nucleic
CC      acids encoding them, vectors, antibodies, host cells which express
CC      heterologous osteoregulins, and animal cells and mammals with a
CC      targeted disruption of an osteoregulin gene. The invention also
CC      provides screening assays to identify modulators of osteoregulin
CC      activity as well as methods of treating mammals for diseases or
CC      disorders associated with osteoregulin activity. The modulators of
CC      activity may be useful in the manufacture of a medicament for, as
CC      well as for treating, a mammal in need of regulation of bone mass
CC      and/or density, adiposity, vascular flexibility, and/or
CC      atherosclerotic plaque calcification (claimed), for treating and
CC      preventing osteoporosis, and for stimulating bone repair and
CC      regeneration.
XX      Sequence 1969 BP; 732 A; 353 C; 447 G; 437 T; 0 other;
SQ      Query Match          99.0%; Score 1638; DB 22; Length 1969;
        Best local Similarity 100.0%; Pred. No. 0;
        Matches 1638; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      2 TGAATAAAGATATAGTATCAGTAACAAAGAGATATCTCAGATGGCCCTGAGATGTCAA 61
DB      332 TGAATAAAGATATAGTATCAGTAACAAAGAGATATCTCAGATGGCCCTGAGATGTCAA 391
QY      62 TTTATCTAGTCACTCGGAATTAAGGGTTTGAGATGAGATGATGATGATGAGCAAAAC 121
DB      392 TTTATCTAGTCACTCGGAATTAAGGGTTTGAGATGAGATGATGATGATGATGAGCAAAAC 451
QY      122 TATATGACCAAGAAGATATGAGGAGCTCTCATCGAATAATACATGCAATATATATG 181
DB      452 TATATGACCAAGAAGATATGAGGAGCTCTCATCGAATAATACATGCAATATATATG 511
QY      182 GGGCAGTACTGCGATTTAACTCTGGGGGAGAGAAAACAAAGAGAACACACTAGGAATG 241
DB      512 GGGCAGTACTGCGATTTAACTCTGGGGGAGAGAAAACAAAGAGAACACACTAGGAATG 571
QY      242 TTCTAAACATTAATCCAGCAAGATATGATATCTTAAGACACCTCGAAGGATTAAGAA 301
DB      572 TTCTAAACATTAATCCAGCAAGATATGATATCTTAAGACACCTCGAAGGATTAAGAA 631
QY      302 AGCCTCAAGAGATTCACCAAGCCAGAAAAGTCCAGTAAAAAACCAAGACCCATCGTA 361

```

```

DB      632 AGCCTCAAGAGATTCACCAAGCCAGAAAAGTCCAGTAAAAAACCAAGACCCATCGTA 691
QY      362 TTCAACACAACTTACTACTTAAACATCTCTCAAAAAGTCAAAAAATCCCGAGTAT 421
DB      692 TTCAACACAACTTACTACTTAAACATCTCTCTCAAAAAGTCAAAAAATCCCGAGTAT 751
QY      422 TTGAAGGACGCGTTATACAGATCTTCAAGAGAGAGGGGCAATGATATATCTCTTCA 481
DB      752 TTGAAGGACGCGTTATACAGATCTTCAAGAGAGAGGGGCAATGATATATCTCTTCA 811
QY      482 GTGGGACGGCCCAACTTTTAAGGACATTCCTGGTAAAGAGAAAGTACGTCCTGACC 541
DB      812 GTGGGACGGCCCAACTTTTAAGGACATTCCTGGTAAAGAGAAAGTACGTCCTGACC 871
QY      542 TAGAAGCAAAAGATATTTCAACAGGGTTTGCAGGCCCACTGAAGCTGAGATCTATC 601
DB      872 TAGAAGCAAAAGATATTTCAACAGGGTTTGCAGGCCCACTGAAGCTGAGATCTATC 931
QY      602 TTGACACAAAAGAGCCAGGTTATATAGATTCCTCAGAGAGAGAGAAATGTTGAATA 661
DB      932 TTGACACAAAAGAGCCAGGTTATATAGATTCCTCAGAGAGAGAGAAATGTTGAATA 991
QY      662 CCAATGGAACTAGGATGAAGAACTGCGAAAGAGGACAGATGCTGTGATGTCAGCTGTAG 721
DB      992 CCAATGGAACTAGGATGAAGAACTGCGAAAGAGGACAGATGCTGTGATGTCAGCTGTAG 1051
QY      722 AGGGCAGCAACGATATCATGAGTATGATCAACATTTTAAGAGCTCCCTGGAAGAGAA 781
DB      1052 AGGGCAGCAACGATATCATGAGTATGATCAACATTTTAAGAGCTCCCTGGAAGAGAA 1111
QY      782 ACAGAGTGGATGGTGGCAGCCAAATGCTCACCAGAGGAGTTGATTCATTCACCTC 841
DB      1112 ACAGAGTGGATGGTGGCAGCCAAATGCTCACCAGAGGAGTTGATTCATTCACCTC 1171
QY      842 CTCACCCCTCAAAAAGAGAAAAGAGAGGAGGAGTGTGATGATGATGATGATGATGAT 901
DB      1172 CTCACCCCTCAAAAAGAGAAAAGAGAGGAGGAGTGTGATGATGATGATGATGATGAT 1231
QY      902 ATATATATATCTTAAATATGCAAAAGGAGGATGATGATGATGATGATGATGATGAT 961
DB      1232 ATATATATATCTTAAATATGCAAAAGGAGGATGATGATGATGATGATGATGATGAT 1291
QY      962 GGAACCAAGCACTTAAATATGAAAACAAAGGTTCTGTATGAGGCAAAAGCAGGGCC 1021
DB      1292 GGAACCAAGCACTTAAATATGAAAACAAAGGTTCTGTATGAGGCAAAAGCAGGGCC 1351
QY      1022 TGCCCATTCCTCTCTGCTGCTTTGATATGAATCAAAAACGAATGATTCCTTTAATG 1081
DB      1352 TGCCCATTCCTCTCTGCTGCTTTGATATGAATCAAAAACGAATGATTCCTTTAATG 1411
QY      1082 GCGCCAGTATGAGATATATATATACATGCGCAAAATATATATATATATATATATAT 1141
DB      1412 GCGCCAGTATGAGATATATATATATACATGCGCAAAATATATATATATATATATAT 1471
QY      1142 AAAATATATCTACAGGATTAAGGATGATGCAAGAGGAAAGGCTCGGGGTAGCAAC 1201
DB      1472 AAAATATATCTACAGGATTAAGGATGATGCAAGAGGAAAGGCTCGGGGTAGCAAC 1531
QY      1202 CCCATTCAACAGAGGATTTAGTCCCGTAGAAGGATGACAGTATGATCATCTGACA 1261
DB      1532 CCCATTCAACAGAGGATTTAGTCCCGTAGAAGGATGACAGTATGATCATCTGACA 1591
QY      1262 GTGGGAGTCAAGTGAAGAGGATGGTACTATCCACAGAGGATTCACAGCGGGGTGACA 1321
DB      1592 GTGGGAGTCAAGTGAAGAGGATGGTACTATCCACAGAGGATTCACAGCGGGGTGACA 1651
QY      1322 GTCTGAAGACCTGTCACCTGTGATGATGATGATGATGATGATGATGATGATGATGAT 1381
DB      1652 GTCTGAAGACCTGTCACCTGTGATGATGATGATGATGATGATGATGATGATGATGAT 1711
QY      1382 GTGAAGAGAGATATAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1441
DB      1712 GTGAAGAGAGATATAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1771

```


QY 242 TTCTAATCATATCCAGCAAGATGATGATTAAGTCTAAAGCACACTCGAAGGATAAAAAGA 301
 |||||
 Db 715 TTCTAATCATATCCAGCAAGATGATGATTAAGTCTAAAGCACACTCGAAGGATAAAAAGA 774
 QY 302 AGCCTCAAAAGATTTCCCAAGCCAGAAAGTCCAGTAAAAAGCAAAAGCCCATCGTA 361
 |||||
 Db 775 AGCCTCAAAAGATTTCCCAAGCCAGAAAGTCCAGTAAAAAGCAAAAGCCCATCGTA 834
 QY 362 TTCAACACACATTTGATCTACCTTAATAATCTCTCAAAAGTCAAAAATATCCCATGAT 421
 |||||
 Db 835 TTCAACACACATTTGATCTACCTTAATAATCTCTCAAAAGTCAAAAATATCCCATGAT 894
 QY 422 TTGAAGCAGCGGTTATACAGATCTTCAAGAGAGAGGGACATGATATATCTCTTCA 481
 |||||
 Db 895 TTGAAGCAGCGGTTATACAGATCTTCAAGAGAGAGGGACATGATATATCTCTTCA 954
 QY 482 GTGGGAGCGGCAACCTTTTAAGGACATCTCTGTTAAAGGAGAGTACTGCTGCTGAC 541
 |||||
 Db 955 GTGGGAGCGGCAACCTTTTAAGGACATCTCTGTTAAAGGAGAGTACTGCTGCTGAC 1014
 QY 542 TAGAAGCAAAAGATATTCAAACAGGGTTTGAGGCCCAAGTGAAGCTGAGATCTATC 601
 |||||
 Db 1015 TAGAAGCAAAAGATATTCAAACAGGGTTTGAGGCCCAAGTGAAGCTGAGATCTATC 1074
 QY 602 TTGACACAAAAAGCCAGGTTATATAGATCCCAAGAGAGAGAAATGTTGGAATA 661
 |||||
 Db 1075 TTGACACAAAAAGCCAGGTTATATAGATCCCAAGAGAGAGAAATGTTGGAATA 1134
 QY 662 CCATTGGAAGTAGGAGTGAAGTCTGGAAGAGAGAGATGCTGATGTCAGCTTGAG 721
 |||||
 Db 1135 CCATTGGAAGTAGGAGTGAAGTCTGGAAGAGAGAGATGCTGATGTCAGCTTGAG 1194
 QY 722 AGGCGACCAAGATATCATGGGTAGTACCAATTTTAAGAGACCTCCCTGGAAGAGAGAA 781
 |||||
 Db 1195 AGGCGACCAAGATATCATGGGTAGTACCAATTTTAAGAGACCTCCCTGGAAGAGAGAA 1254
 QY 782 ACAGAGTGTGCTGCGAGCCAAATGCTCAGCCAGGAGGTTGATTTCTATACCTC 841
 |||||
 Db 1255 ACAGAGTGTGCTGCGAGCCAAATGCTCAGCCAGGAGGTTGATTTCTATACCTC 1314
 QY 842 CTGCACCTCTCAAAAGAAAAAGAAAGAGAGAGTATGATGATGATGATGATGAT 901
 |||||
 Db 1315 CTGCACCTCTCAAAAGAAAAAGAAAGAGAGAGTATGATGATGATGATGATGAT 1374
 QY 902 ATATATGAATTTCTAAAAATGCGAAGGAGTACAGAAAGGTTGATGATGATGAT 961
 |||||
 Db 1375 ATATATGAATTTCTAAAAATGCGAAGGAGTACAGAAAGGTTGATGATGATGAT 1434
 QY 962 GGAACCAAGAGACCTTAATGAAGAAAAAAGGTTCTAGTAAAGGCAAAAGTACAGGCC 1021
 |||||
 Db 1435 GGAACCAAGAGACCTTAATGAAGAAAAAAGGTTCTAGTAAAGGCAAAAGTACAGGCC 1494
 QY 1022 TGCCCATTTCTCTGCTGCTGCTGATGATGATGATGATGATGATGATGATGATG 1081
 |||||
 Db 1495 TGCCCATTTCTCTGCTGCTGCTGATGATGATGATGATGATGATGATGATGATG 1554
 QY 1082 GCGCCAGTATGAGATATATATTAACACATGCGCAGAAAAATATCATTTATGCCACAGAC 1141
 |||||
 Db 1555 GCGCCAGTATGAGATATATATTAACACATGCGCAGAAAAATATCATTTATGCCACAGAC 1614
 QY 1142 AAAATATTTCTACACGGAATTAAGGTTATGCCACAGGAAAAAGGTTCTGGGGTACACAC 1201
 |||||
 Db 1615 AAAATATTTCTACACGGAATTAAGGTTATGCCACAGGAAAAAGGTTCTGGGGTACACAC 1674
 QY 1202 CCCATTCAACAGAGGTTTATGTTCCGTTAGAAAGGATGACAGTAGTAGTCACTGAC 1261
 |||||
 Db 1675 CCCATTCAACAGAGGTTTATGTTCCGTTAGAAAGGATGACAGTAGTAGTCACTGAC 1734
 QY 1262 GTGGCAGTTCAAGTGAAGAGGATGATGATGATGATGATGATGATGATGATGATG 1321
 |||||
 Db 1735 GTGGCAGTTCAAGTGAAGAGGATGATGATGATGATGATGATGATGATGATGATG 1794
 QY 1322 GTCTGAAGACCTGCTACCTGTGTGATGATGATGATGATGATGATGATGATGATG 1381

Db 1795 GTCTGAAGACCTGCTACCTGTGATGATGATGATGATGATGATGATGATGATGATG 1854
 QY 1382 GTGAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1441
 |||||
 Db 1855 GTGAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1914
 QY 1442 TTTTGTCTATCTTAATAGTCACTATTAATAATCTTATTAAGGCTATATGTTTAAAGC 1501
 |||||
 Db 1915 TTTTGTCTATCTTAATAGTCACTATTAATAATCTTATTAAGGCTATATGTTTAAAGC 1974
 QY 1502 AAAAAAATATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1561
 |||||
 Db 1975 AAAAAAATATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2034
 QY 1562 TAGTTGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1621
 |||||
 Db 2035 TAGTTGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2094
 QY 1622 AATATCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1639
 |||||
 Db 2095 AATATCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2112
 |||||
 RESULT 8
 AAK94437 standard; cDNA: 2031 BP.
 ID AAK94437;
 AC AAK94437;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human full-length cDNA, SEQ ID NO: 3222.
 XX
 KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
 XX
 OS Homo sapiens.
 XX
 PN EP1130094-A2.
 PD
 XX 05-SEP-2001.
 PE 07-JUL-2000; 2000EP-0114089.
 XX
 PR 08-JUL-1999; 99JP-0194486.
 PR 11-JAN-2000; 2000JP-0118774.
 PR 02-MAY-2000; 2000JP-0183765.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 XX
 DR WPI: 2001-524255/58.
 DR P-PSDB: AAM93507.
 PT
 PT 830 Primers useful for synthesizing full length cDNA clones and their
 use in genetic manipulation -
 PS Claim 8; SEQ ID NO 3222; 1380bp + sequence listing; English.
 CC The invention relates to primers for synthesizing full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been
 CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
 CC molecules have been determined. Primers for synthesizing the full length
 CC cDNA are useful for clarifying the function of the protein encoded by
 CC the cDNA. The full length clones were obtained by construction of full
 CC length enriched cDNA libraries that were synthesised by the oligo-capping
 CC method. The primers enable the production of the full length cDNA easily
 CC without any special methods. The present sequence is a full length
 CC human cDNA of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in CD-ROM format directly from EPO.

XX Sequence 2031 BP; 740 A; 374 C; 463 G; 454 T; 0 other:

Query Match 98.2%; Score 1624.4; DB 22; Length 2031;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 1636; Conservative 0; Mismatches 1; Indels 1; Gaps 1.

```

OY 2 TGAATTAAGATATAGTATCAGTAACAAAGAGAAATCTCACANTGGCCGAGAGTGCAA 61
DB 373 TGAATTAAGATATAGTATCAGTAACAAAGAGAAATCTCACANTGGCCGAGAGTGCAA 432
OY 62 TTTATCTTAAGTCAACTGGGAATAAAGGTTTGAGGATGAGATGATGCTATCAGCAAC 121
DB 433 TTTATCTTAAGTCAACTGGGAATAAAGGTTTGAGGATGAGATGATGCTATCAGCAAC 492
OY 122 TACATGACCAAGAGATATGGCGAGCTCTCATCAGAAATTAACATGCAACATATATG 181
DB 493 TACATGACCAAGAGATATGGCGAGCTCTCATCAGAAATTAACATGCAACATATATG 552
OY 182 GGCCAGTGCATCGATTAACCTCTGGGGGAGAAACAAAGAGAACACACCTAGGAATG 241
DB 553 GGCCAGTGCATCGATTAACCTCTGGGGGAGAAACAAAGAGAACACACCTAGGAATG 612
OY 242 TTCTAAACATTAATCCAGCAAGTATGTAATGCTTAAGCAGCTCGAAGATTAATAA 301
DB 613 TTCTAAACATTAATCCAGCAAGTATGTAATGCTTAAGCAGCTCGAAGATTAATAA 672
OY 302 AGCTTCAGAGATTCCTCCAGCCAGAAAGATCGCTTAAGCAAAAGCACCACCTGTA 361
DB 673 AGCTTCAGAGATTCCTCCAGCCAGAAAGATCGCTTAAGCAAAAGCACCACCTGTA 732
OY 362 TTCAACACAAATGCTACTCTAAACATCTCTCAAAAGTCAAAAATCCCAAGTAT 421
DB 733 TTCAACACAAATGCTACTCTAAACATCTCTCAAAAGTCAAAAATCCCAAGTAT 792
OY 422 TTGAAGCGACGGTTATACAGATCTTCAAGAGAGAGGAGACATGATATATCTCTTCA 481
DB 793 TTGAAGCGACGGTTATACAGATCTTCAAGAGAGAGGAGACATGATATATCTCTTCA 852
OY 482 GTGGGAGCGCCACCTTTAAGGACATCTCTGTTAAAGAGAGAGTACTGCTCGAC 541
DB 853 GTGGGAGCGCCACCTTTAAGGACATCTCTGTTAAAGAGAGAGTACTGCTCGAC 911
OY 542 TAGAAGCGAAAGATATTCACAGGGTTTGCAGGCCCAAGTGAAGCTGAGACTCATC 601
DB 912 TAGAAGCGAAAGATATTCACAGGGTTTGCAGGCCCAAGTGAAGCTGAGACTCATC 971
OY 602 TTGACACAAAAGCCAGGTTATATGAGATCCAGAGAGAGAGAAATGGTGAATA 661
DB 972 TTGACACAAAAGCCAGGTTATATGAGATCCAGAGAGAGAGAAATGGTGAATA 1031
OY 662 CCATTGGAATAGGATGAAATCGGAAAGAGGAGATCTGTTGATGTCACCTTGTG 721
DB 1032 CCATTGGAATAGGATGAAATCGGAAAGAGGAGATCTGTTGATGTCACCTTGTG 1091
OY 722 AGGCGAGCAACATATCATGAGTATGACCAATTTTAAGAGTCCCTGGAAGAGAA 781
DB 1092 AGGCGAGCAACATATCATGAGTATGACCAATTTTAAGAGTCCCTGGAAGAGAA 1151
OY 782 ACAGAGTGAATCTGGCAGCCAAATGCTCACCAAGGAGGTTGATTTCAATTACCTC 841
DB 1152 ACAGAGTGAATCTGGCAGCCAAATGCTCACCAAGGAGGTTGATTTCAATTACCTC 1211
OY 842 CTGCAACCCCTCAAAAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 901
DB 1212 CTGCAACCCCTCAAAAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1271
OY 902 ATATGAATTTCTAAATATGCAAAAGGAGTACCAAGAGAGGTTGATGATCTTATA 961
DB 1272 ATATGAATTTCTAAATATGCAAAAGGAGTACCAAGAGAGGTTGATGATCTTATA 1331
OY 962 GGAACCAAGCACTTAAATGAAAAAAGGTTTCTGTAAGGCAAAAGCTCAGGCC 1021
DB 1332 GGAACCAAGCACTTAAATGAAAAAAGGTTTCTGTAAGGCAAAAGCTCAGGCC 1391
OY 1022 TGGCCATTCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1081
DB 1392 TGGCCATTCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1451
OY 1082 GCCCAGTATGAGATATTAATTAACATGAGAGAGAGAGAGAGAGAGAGAGAGAG 1141
DB 1452 GCCCAGTATGAGATATTAATTAACATGAGAGAGAGAGAGAGAGAGAGAGAGAG 1511
OY 1142 AAAATTAATTTCTACAGGAAATAGGTTATGCCAAGAGAGAGAGAGAGAGAGAG 1201
DB 1512 AAAATTAATTTCTACAGGAAATAGGTTATGCCAAGAGAGAGAGAGAGAGAGAG 1571
OY 1202 CCCATTCCAGAGAGAGGTTTACTTCCCGTAGAAGAGAGATGACAGTATGATCGACA 1261
DB 1572 CCCATTCCAGAGAGAGGTTTACTTCCCGTAGAAGAGAGATGACAGTATGATCGACA 1631
OY 1262 GTGGCAGTTCAAGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1321
DB 1632 GTGGCAGTTCAAGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1691
OY 1322 GTCTGAAGACCTGTCACCTGAGAGTATGATGATGATGATGATGATGATGATG 1381
DB 1692 GTCTGAAGACCTGTCACCTGAGAGTATGATGATGATGATGATGATGATGATG 1751
OY 1382 GTGAAGAGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1441
DB 1752 GTGAAGAGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1811
OY 1442 TTTTGTCTATTTAAATGACAGTATTAATTAATTAATTAATTAATTAATTAAT 1501
DB 1812 TTTTGTCTATTTAAATGACAGTATTAATTAATTAATTAATTAATTAATTAAT 1871
OY 1502 AAAAAAAATTCATACAGATCTATGAATAGTAAATGATTAATGATGATGATTA 1561
DB 1872 AAAAAAAATTCATACAGATCTATGAATAGTAAATGATTAATGATGATGATTA 1931
OY 1562 TAGTTGAGTAAATGTCACAAATGCTTCTGTTGTTGTTGTTGTTGTTGTTGTT 1621
DB 1932 TAGTTGAGTAAATGTCACAAATGCTTCTGTTGTTGTTGTTGTTGTTGTTGTT 1691
OY 1622 AATATCTCTGATGATTA 1639
DB 1992 AATATCTCTGATGATTA 2009

```

```

DB 1332 GGAACCAAGCACTTAAATGAAAAAAGGTTTCTGTAAGGCAAAAGCTCAGGCC 1391
OY 1022 TGGCCATTCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1081
DB 1392 TGGCCATTCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1451
OY 1082 GCCCAGTATGAGATATTAATTAACATGAGAGAGAGAGAGAGAGAGAGAGAGAG 1141
DB 1452 GCCCAGTATGAGATATTAATTAACATGAGAGAGAGAGAGAGAGAGAGAGAGAG 1511
OY 1142 AAAATTAATTTCTACAGGAAATAGGTTATGCCAAGAGAGAGAGAGAGAGAGAG 1201
DB 1512 AAAATTAATTTCTACAGGAAATAGGTTATGCCAAGAGAGAGAGAGAGAGAGAG 1571
OY 1202 CCCATTCCAGAGAGGTTTACTTCCCGTAGAAGAGAGATGACAGTATGATCGACA 1261
DB 1572 CCCATTCCAGAGAGAGGTTTACTTCCCGTAGAAGAGAGATGACAGTATGATCGACA 1631
OY 1262 GTGGCAGTTCAAGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1321
DB 1632 GTGGCAGTTCAAGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1691
OY 1322 GTCTGAAGACCTGTCACCTGAGAGTATGATGATGATGATGATGATGATGATG 1381
DB 1692 GTCTGAAGACCTGTCACCTGAGAGTATGATGATGATGATGATGATGATGATG 1751
OY 1382 GTGAAGAGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1441
DB 1752 GTGAAGAGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1811
OY 1442 TTTTGTCTATTTAAATGACAGTATTAATTAATTAATTAATTAATTAATTAAT 1501
DB 1812 TTTTGTCTATTTAAATGACAGTATTAATTAATTAATTAATTAATTAATTAAT 1871
OY 1502 AAAAAAAATTCATACAGATCTATGAATAGTAAATGATTAATGATGATGATTA 1561
DB 1872 AAAAAAAATTCATACAGATCTATGAATAGTAAATGATTAATGATGATGATTA 1931
OY 1562 TAGTTGAGTAAATGTCACAAATGCTTCTGTTGTTGTTGTTGTTGTTGTTGTT 1621
DB 1932 TAGTTGAGTAAATGTCACAAATGCTTCTGTTGTTGTTGTTGTTGTTGTTGTT 1691
OY 1622 AATATCTCTGATGATTA 1639
DB 1992 AATATCTCTGATGATTA 2009

RESULT 9
ABA99160
ID ABA99160 standard; DNA; 1662 BP.
XX
AC ABA99160;
XX
DE 23-MAY-2002 (first entry)
XX
DE Human phosphatoin encoding sequence including untranslated region.
XX
KW Human phosphatoin; cytosolic; antidiabetic; antinflammatory;
KW hyperphosphatemia; arteriosclerosis; heart failure; gene;
KW diabetic renal disease; kidney failure; cystic fibrosis; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 8..1585
FT FT /*tag= a
FT FT /product= "Human phosphatoin"
PN WO200198495-A1.
PD 27-DEC-2001.
XX
PF 20-JUN-2001; 2001WO-JP05263.

```

XX PR 21-JUN-2000; 2000JP-0191088.
XX PA (TAKE) TAKEDA CHEM IND LTD.
XX PI
XX Kurokawa T, Yamada T, Morimoto S;
DR WPI; 2002-139791/18.
DR P-PSTDB; ABB08526.

XX PT Phosphatonin of human origin and DNA encoding it for diagnosis and
PT treatment of diseases associated with disorders of phosphate
PT metabolism, e.g., hyperphosphatemia, arteriosclerosis, heart failure,
PS diabetic renal disease and kidney failure - Japanese.
PS Claim 9; Page 123-124; 130pp; Japanese.

XX CC This invention relates to human phosphatonin which has
CC the activity of lowering blood phosphate and increasing urinary
CC phosphate. The proteins are cytosolic, antidiabetic and
CC antiinflammatory in their action. Phosphatonin down-regulates
CC sodium-dependent phosphate transport in the kidney, it down-regulates
CC 25-hydroxy-vitamin D3-1alpha-hydroxylase in the kidney and up-regulates
CC 25-hydroxy-vitamin D3-24-hydroxylase in the kidney. Phosphatonin is
CC useful in the diagnosis, treatment and prevention of phosphate
CC metabolism related diseases such as hyperphosphatemia, arteriosclerosis,
CC heart failure, diabetic renal disease, kidney failure, acute coronary
CC disease and cystic fibrosis. This sequence represents human
CC phosphatonin encoding sequence and includes untranslated regions.

XX Sequence 1662 BP; 610 A; 317 C; 388 G; 347 T; 0 other;

Query Match	82.7%;	Score 1369;	DB 24;	Length 1662;
Best Local Similarity	100.0%;	Pred. No. 0;		

Matches 1369; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY	2	TGATTAAGGATPATAGTATCGTATACAAAGGATPACTACAAATGGCTGAGATGTCAA	61
Db	294	TGAATTAAGGATPATAGTATCGTATACAAAGGATPACTACAAATGGCTGAGATGTCAA	35
OY	62	TTTATCTTAAGTCAACTGGGAATTAAGGGTTTGAGGATGGAGATGATGATCGATCGCAAC	121
Db	354	TTTATCTTAAGTCAACTGGGAATTAAGGGTTTGAGGATGGAGATGATGATCGATCGCAAC	413
OY	122	TACATGACCAAGAGATATGGCGAGCTCTCATCAGAAATACATCGAACATTAATGG	181
Db	414	TACATGACCAAGAGATATGGCGAGCTCTCATCAGAAATACATCGAACATTAATGG	473
OY	182	GGCCAGTACAGCGATTAACTCCGCGGGGGAAGAAACAAAGAACACACACTGAGATG	241
Db	474	GGCCAGTACAGCGATTAACTCCGCGGGGGAAGAAACAAAGAACACACACTGAGATG	533
OY	242	TTCTTAACATATATCCACAGCATATGAATTTATGCTAAAGCACATCGAAGATTAAGA	301
Db	534	TTCTTAACATATATCCACAGCATATGAATTTATGCTAAAGCACATCGAAGATTAAGA	593
OY	302	AGCCTCAAGAGATTTCCCAAGCCACAGAAAGTCCAGTAAAAAGCAAGCACCATCGTA	361
Db	594	AGCCTCAAGAGATTTCCCAAGCCACAGAAAGTCCAGTAAAAAGCAAGCACCATCGTA	653
OY	362	TTCAACACAAATGACACTACTAAACACTCTCTAAAAAGTCAAAAAATCCCGATATT	421
Db	654	TTCAACACAAATGACACTACTAAACACTCTCTAAAAAGTCAAAAAATCCCGATATT	713
OY	422	TTGAAGGACGGTTATCTACATCTTCAAGAGAGAGGGGACAAATGATATATCTCTTCA	481
Db	714	TTGAAGGACGGTTATCTACATCTTCAAGAGAGAGGGGACAAATGATATATCTCTTCA	773
OY	482	GTGGGAGCGCCAACTTTTAAGACATTTCTGTAAAGGAAGAGTACTGTCCTGACC	541
Db	774	GTGGGAGCGCCAACTTTTAAGACATTTCTGTAAAGGAAGAGTACTGTCCTGACC	833
OY	542	TAGAAGGCAAAATATTCAACAGGGTTTGACAGGCCCAAGTGAAGCTGAGATCTCATC	601

Db	834	TAGAAGCAAGAATATTCAACACAGGGTTTGACAGGCCCAAGTAGAAGCTGAGAGTACTCATC	893
Oy	602	TTGACACAAAAAAGCCAGGTTATATATGAGATCCGACAGAGAGAAAGAAAATGGTGAATA	661
Db	894	TTGACACAAAAAAGCCAGGTTATATATGAGATCCGACAGAGAGAAAGAAAATGGTGAATA	953
Oy	662	CCATTGGACACTAGGAGTGAAGAACTGCCAAAAGAGACATCTCTTGATGTCAAGCTTGTAG	721
Db	954	CCATTGGAACTAGGAGTGAAGAACTGCCAAAAGAGACATCTCTTGATGTCAAGCTTGTAG	1013
Oy	722	AGGCGACCAACGATATCATCGGTAGTACCAATTTTAAGSAGCTCCCTGGAAAGAGAA	781
Db	1014	AGGCGACCAACGATATCATCGGTAGTACCAATTTTAAGSAGCTCCCTGGAAAGAGAA	1073
Oy	782	ACAGATGGATGCTGCGCAGCCAAAATGCTCCACAGAGGAGTTGAGTTCTTACCTC	841
Db	1074	ACAGATGGATGCTGCGCAGCCAAAATGCTCCACAGAGGAGTTGAGTTCTTACCTC	1133
Oy	842	CTGCAACCCCTCAAAAAGAGAAAAGAAAGGCAAGTGTATGATCAGCTGAAGTACCACAT	901
Db	1134	CTGCAACCCCTCAAAAAGAGAAAAGAAAGGCAAGTGTATGATCAGCTGAAGTACCACAT	1193
Oy	902	ATAATGAATTCCTAAAAATGGCAAGGCAAGTACAGAAAGGGTTAGATCTTTAATA	961
Db	1194	ATAATGAATTCCTAAAAATGGCAAGGCAAGTACAGAAAGGGTTAGATCTTTAATA	1253
Oy	962	GGAAACCAAGCAACCTTAATGAAAACAAAGGTTTCTGTAGTAAGGGCAAAAGCTCAGGGCC	1021
Db	1254	GGAAACCAAGCAACCTTAATGAAAACAAAGGTTTCTGTAGTAAGGGCAAAAGCTCAGGGCC	1313
Oy	1022	TGCCCCATTCCTCTCTCGTGCTCTTGATATATGAAATCAAAAAAGCAATGGATTCCTTAATG	1081
Db	1314	TGCCCCATTCCTCTCTCGTGCTCTTGATATATGAAATCAAAAAAGCAATGGATTCCTTAATG	1373
Oy	1082	GCCCCAGTCATGAGAATATATATTAACACATGGCGAGAAATATCATTTATGATGACCCACAGAC	1141
Db	1374	GCCCCAGTCATGAGAATATATATTAACACATGGCGAGAAATATCATTTATGATGACCCACAGAC	1433
Oy	1142	AAAAATAATTCACACAGGAATAGGGTATGCGCACAGAGAAAGGCTCTGGGGTAGACAAC	1201
Db	1434	AAAAATAATTCACACAGGAATAGGGTATGCGCACAGAGAAAGGCTCTGGGGTAGACAAC	1493
Oy	1202	CCCATTCACACAGGAGGTTTAGTTCCCGTATGAAGGATACAGTACTGATCATCTGACA	1261
Db	1494	CCCATTCACACAGGAGGTTTAGTTCCCGTATGAAGGAGTACAGTACTGATCATCTGACA	1553
Oy	1262	GTGGCAGTTCACAGTACAGGAGGATGGTACATGCCACAGGAGTTCCACGGGGGTGACA	1321
Db	1554	GTGGCAGTTCACAGTACAGGAGGATGGTACATGCCACAGGAGTTCCACGGGGGTGACA	1613
Oy	1322	GTCTGAGAGACCTGCTCACTGTTGAGTTGATGTAGAGAGAGGCCACCTGA	1370
Db	1614	GTCTGAGAGACCTGCTCACTGTTGAGTTGATGTAGAGAGAGGCCACCTGA	1662
RESULT 10			
ABA99159			
AC	ABA99159	standard; DNA; 1575 BP.	
XX	23-MAY-2002	(first entry).	
DE	Human phosphatoin encoding sequence.		
KW	Human phosphatoin; cytosolic; antidiabetic; antiinflammatory;		
KW	hyperphosphemia; arteriosclerosis; heart failure; gene;		
KW	diabetic renal disease; kidney failure; cystic fibrosis;		
XX	phosphate transport; ds.		
OS	Homo sapiens.		
XX			

FH Key Location/Qualifiers
 FT CDS 1..1575
 FT /*tag- a
 FT /partial
 FT /note="no stop codon"
 FT /product="human phosphatonin"
 PN MO200198495-A1.
 XX
 XX 27-DEC-2001.
 XX
 PF 20-JUN-2001; 2001MO-JP05263.
 XX
 PR 21-JUN-2000; 2000JP-0191088.
 XX
 PA (TAKE) TAKEDA CHEM IND LTD.
 XX
 XX Kurokawa T, Yamada T, Morimoto S;
 PI WPI; 2002-139791/18.
 DR P-PSDB; ABB08526.
 XX
 XX Phosphatonin of human origin and DNA encoding it for diagnosis and
 PT treatment of diseases associated with disorders of phosphate
 PT metabolism, e.g., hyperphosphatemia, arteriosclerosis, heart failure,
 PT diabetic renal disease and kidney failure
 XX
 XX Claim 9; Fig 1-4; 130pp; Japanese.
 XX
 CC This invention relates to human phosphatonin which has
 CC the activity of lowering blood phosphate and increasing urinary
 CC phosphate. The proteins are cytosolic, antidiabetic and
 CC antiinflammatory in their action. Phosphatonin down-regulates
 CC sodium-dependent phosphate transport in the kidney, it down-regulates
 CC 25-hydroxy-vitamin D3-1alpha-hydroxylase in the kidney and up-regulates
 CC 25-hydroxy-vitamin D3-24-hydroxylase in the kidney. Phosphatonin is
 CC useful in the diagnosis, treatment and prevention of phosphate
 CC metabolism related diseases such as hyperphosphatemia, arteriosclerosis,
 CC heart failure, diabetic renal disease, kidney failure, acute coronary
 CC disease and cystic fibrosis. This sequence represents human
 CC phosphatonin encoding sequence. The full sequence including
 CC untranslated regions is given in ABA99160.
 XX
 XX Sequence 1575 BP; 588 A; 295 C; 362 G; 330 T; 0 other:
 SQ
 Query Match 77.9%; Score 1289; DB 24; Length 1575;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1289; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 TTGACACAAACATTGACTACTTAAACATCTCTCAAAAGTCAAAAAATCCCGATGAT 421
 DB 647 TTCAACCAACATGACTACTACTTAAACATCTCTCAAAAGTCAAAAAATCCCGATGAT 706
 QY 422 TTGAAGGACGCGGTATATACAGATCTTCAAGAGAGAGGGGCAATGATATATCTCTTCA 481
 DB 707 TTGAAGGACGCGGTATATACAGATCTTCAAGAGAGAGGGGCAATGATATATCTCTTCA 766
 QY 482 GTGGGACGCGCAACCTTTTAAGACATCTCTGTTAAAGAGAGAAAGCTACTGTCGAC 541
 DB 767 GTGGGACGCGCAACCTTTTAAGACATCTCTGTTAAAGAGAGAAAGCTACTGTCGAC 826
 QY 542 TTGAAGGACAAAGATATTTCAACAGGTTTGCAGGCCCAAGTGAAGTGAAGTACTATC 601
 DB 827 TTGAAGGACAAAGATATTTCAACAGGTTTGCAGGCCCAAGTGAAGTGAAGTACTATC 886
 QY 602 TTGACACAAAAAAAGCAGGTTATATAGATATCCAGAGAGAGAAAGAAATGGTGGAAATA 661
 DB 887 TTGACACAAAAAAAGCAGGTTATATAGATATCCAGAGAGAGAAAGAAATGGTGGAAATA 946
 QY 662 CCATTGGAACTAGGATGTAACCTGCGAAAAGGACAGATCTGTTGATGTCAGCCTTTGAG 721
 DB 947 CCATTGGAACTAGGATGTAACCTGCGAAAAGGACAGATCTGTTGATGTCAGCCTTTGAG 1006
 QY 722 AGGGCAGCAACGATATCATGAGTACCAATTTTAAGAGAGTCCCTGGAAAGAAAGGA 781
 DB 1007 AGGGCAGCAACGATATCATGAGTACCAATTTTAAGAGAGTCCCTGGAAAGAAAGGA 1066
 QY 782 ACAGAGTGGATCTGGCAGCCAAAATGCTCACCAAGGAGGTTGATGTTCTTACCTTC 841
 DB 1067 ACAGAGTGGATCTGGCAGCCAAAATGCTCACCAAGGAGGTTGATGTTCTTACCTTC 1126
 QY 842 CAGCAGCCCTCAAAAGAGAAAGAAAGAAAGGACATGATGATCAGCTGAAAGTACCAACT 901
 DB 1127 CAGCAGCCCTCAAAAGAGAGAAAGAAAGGACATGATGATCAGCTGAAAGTACCAACT 1186
 QY 902 ATATATGAATTTCTTAAATATGCAAAAGGACATGACAGAAAGGTTAGATCTTCTAATA 961
 DB 1187 ATATATGAATTTCTTAAATATGCAAAAGGACATGACAGAAAGGTTAGATCTTCTAATA 1246
 QY 962 GGAACCAAGCAACCTTAAATGAAAAAAGGTTTCTAGTAAGGCAAAAGTCAAGGCC 1021
 DB 1247 GGAACCAAGCAACCTTAAATGAAAAAAGGTTTCTAGTAAGGCAAAAGTCAAGGCC 1306
 QY 1022 TGGCATTCTCTCTGCTGCTGTTGATATGAAATCAAAAGCAAAATGATTCCTTAATG 1081
 DB 1307 TGGCATTCTCTCTGCTGCTGTTGATATGAAATCAAAAGCAAAATGATTCCTTAATG 1366
 QY 1082 GCCCAGTCATGAGATATATATATACATGCGAGAGAAATATCATATGACCCAGAC 1141
 DB 1367 GCCCAGTCATGAGATATATATATACATGCGAGAGAAATATCATATGACCCAGAC 1426
 QY 1142 AAAATATATTTCTACAGGAAATAGGTTATGCCAAGGAAAGGCTCTGGGGTACACAAC 1201
 DB 1427 AAAATATATTTCTACAGGAAATAGGTTATGCCAAGGAAAGGCTCTGGGGTACACAAC 1486
 QY 1202 CCCATTCCAACAGAGGTTTGGTCCGTAGAGGAGTGAAGTATGATGATCTGACA 1261
 DB 1487 CCCATTCCAACAGAGGTTTGGTCCGTAGAGGAGTGAAGTATGATGATCTGACA 1546
 QY 1262 GTGGCAGTTCAAGTGAAGCGATGGTGAC 1290
 DB 1547 GTGGCAGTTCAAGTGAAGCGATGGTGAC 1575
 RESULT 11
 AAH26806
 ID AAH26806 standard; cDNA; 1655 BP.
 XX
 XX AAH26806;
 XX
 XX 21-DEC-2001 (first entry)
 XX

DE Rat osteoregulin cDNA.
 XX Osteoregulin: rat; bone; homeostasis; adipose; calcification;
 KM atherosclerosis; osteoporosis; osteopathic; antiarteriosclerotic;
 KW therapy; ss.
 XX Rattus norvegicus.
 OS
 FH Key Location/Qualifiers
 FT CDS 62..1369
 FT sig_peptide /tag= a
 FT 62..109 /tag= b
 FT mat_peptide 110..1366
 FT /tag= c
 XX
 PN EP1130098-A2.
 XX
 PD 05-SEP-2001.
 XX
 PE 27-FEB-2001; 2001EP-0301768.
 XX
 PR 29-FEB-2000; 2000US-185617P.
 PR 22-SEP-2000; 2000US-234500P.
 XX
 PA (PE12) PFIZER PROD INC.
 XX
 PI Brown TA, De Wet JR, Gowen LC, Hames LM;
 XX WPI, 2001-604111/69.
 DR P-PSDB; AAB92818.
 XX
 PT Novel osteoregulin polypeptide useful for regulating bone homeostasis,
 PT adiposity and calcification of atherosclerotic plaques comprises
 PT measuring the activity of osteoregulin -
 XX
 PS Claim 4; Page 40; 90pp; English.
 XX
 CC The present sequence is that of cDNA encoding rat osteoregulin (see
 CC AAB82918), a novel protein which plays a role in regulating bone
 CC homeostasis, adiposity, and the calcification of atherosclerotic
 CC plaques. The cDNA was isolated from a cDNA library enriched for
 CC genes induced during osteoblastic differentiation, which had been
 CC constructed by subtractive hybridisation of cDNA from dexamethasone
 CC treated rat bone marrow cells. The invention provides osteoregulin
 CC proteins, nucleic acids which encode them, vectors, antibodies,
 CC host cells which express heterologous osteoregulin, and animal
 CC cells and mammals with a targeted disruption of an osteoregulin
 CC gene. The invention also provides screening assays to identify
 CC modulators of osteoregulin activity as well as methods of treating
 CC mammals for diseases or disorders associated with osteoregulin
 CC activity. The modulators of activity may be useful in the
 CC manufacture of a medicament for, as well as for treating, a mammal
 CC in need of regulation of bone mass and/or density, adiposity,
 CC vascular flexibility, and/or atherosclerotic plaque calcification
 CC (claimed), for treating and preventing osteoporosis, and for
 CC stimulating bone repair and regeneration.
 CC
 XX Sequence 1655 BP; 503 A; 396 C; 444 G; 312 T; 0 other:
 XX
 SO
 Query Match 31.1%; Score 514.8; DB 22; Length 1655;
 Best Local Similarity 63.0%; Pred. No. 1.9e-121;
 Matches 994; Conservative 0; Mismatches 487; Indels 97; Gaps 9;

Db 290 AAGAGTCTGTGAGCGGAGCTAGTACAGAGGAGACAAAGAGAAAGAACTCAG 349
 Qy 238 AATGTTCTAAACATTAATCCAGCAAGTATGATTAATGCTAAAGCACACTCGAAGATAA 297
 Db 350 AGTGTCTAAGCGTAATTCACAGATGTCACAAATACAGACTATCAGAAAGATACA 409
 Qy 298 AAGAACCTTCAAGAGATTCGCCAAGCGGAAAAATCCAGTAAAAAGCAAAAGCAACCAT 357
 Db 410 GAGAACCAACAGAGAGATCTACTACTCCAGAACAGCCGACAGCAAAAGCAACACCCCT 469
 Qy 358 CGTATTCACACACAACTTGTACTACTTAAACATCTCTCAAAAGTCAAAAAATCCCACT 417
 Db 470 CGGGCCCGAGAGCAAGCACTACCTACATCAACATCTCCCAAAATCAGAAAGATTCTCAGT 529
 Qy 418 GATTTGAAAGCAAGGTTATACAGATCTTCAAGAGAGAGGAGACAAATGATATCTCT 477
 Db 530 GACTTCGAGGACAGTGTCTCCAGACCTTCTAGTAGGGGGATATATGATCTCTCTCT 589
 Qy 478 TTGAGTGGGAGGCGCAACCTTTTAAGACATCTTGTAAGAGAGAGTACTGCTCTCT 537
 Db 590 TTGAGTGGAGATGACACATTTTATGACACTCCGACAGAGAGGTGCTGTGATCT 649
 Qy 538 GACCTAGAAAGCAAAAGATATTCAACAGGGTTTGACAGGCCCAAGTAAAGCTGAGTACT 597
 Db 650 GATCTGAAAGCTCAGCTGCTCACCCTGTGT---CAGGCTCCAGCAATGTCGAGATTGTT 706
 Qy 598 CATCTTGACACAAAAGCCAGGTTAATAGATATCCAGAGAGAGAAATGAGGA 657
 Db 707 GACCCACACAGATGAGCTGCTTAATGAGATCCAGAGGAGAGAGAGTCAATAGGC 766
 Qy 658 AATACCATTTGAACTAGAGATGAACCTGCGAAGAGGACAGATCTGTTGATGACCTT 717
 Db 767 GGTCCCTATGCAACCCAGAAAACCTGCCAGGGGAGCTCCGCGATGTGAGCT 826
 Qy 718 GTAGAGGCGACAAAGATATCATGAGTATCCAAATTTTAAGAGAGCCCTCGAAGAGAA 777
 Db 827 GTGAGGGGCAAGCAATGAAATCAGCGGCAATCAAAATTTAGGGAGCTCCCTGGAAGAA 886
 Qy 778 GGAAGAGAGTGGTGGCGAGCCAAATGCTCACCAGGAGAGGTTGAATTCATTAAC 837
 Db 887 GGAACAGAGTGCATGTCGACAGCAGCCAAATGCTATCAAGAAAGTAGAATTTCACTAC 946
 Qy 838 CCTCTGACCCCTCAAAAGAGAAAAGAAAGAGGAGCTAGTGTGACGTGAAGATACC 897
 Db 947 CCACAAGCGCCTCAAAAGAGAAAGAGAGGGGCGAGGAGCAACAGGGAAGGCC 1006
 Qy 898 AACTATATGAATTCCTTAAATGCAAAAGGAGGAGTCCAGAAAGGCTGTAGATCATCT 957
 Db 1007 GGTATCAATGAATCCCAAGAGCAAGAGGCGGCGCTAGCAAGATGGGGAAGATCT 1066
 Qy 958 AATGAAACCAAGCAACCTTAATGAAAACAAAGGTTTCTAGTAAAGGCAAAAGTACAG 1017
 Db 1067 AAAGGAACCAAGTAACTTCACTGCAAAAGCCAAAGGTTCCAGGCAAAAGGAGGCGCAG 1126
 Qy 1018 GCGCTGCCATCTCTCTCTGCTGCTGATATGAATCAAAAGCAAAAGGATCTCTT 1077
 Db 1127 -----TCTTCTCACAGTCTTGATAGGTTAAAGTGAAGAAAGCTCTTCT 1174
 Qy 1078 AATGGCCCGAGTCATGAGATATATATACACATGCGAGAAATATCATATGATCCCCAC 1137
 Db 1175 AATACTCTCAGTAGAGAGGGATGCAATA-----GCACACAGG 1213
 Qy 1138 AGACAAATAATTTCTACACGGAATAGGGTATGCCAAGAGGAAAGCTCTCTGGG---GT 1194
 Db 1214 AGAACCAACCACTTACACAGGAATAGGGGATGTACACGCGAGAGAGGCTCTGGGCGCTCG 1273
 Qy 1195 AGACACCCCATTTCCAAAGAGAGGTTTATGTTCCGTAAAGAGATAGTATGAGTCA 1254
 Db 1274 AGAAAGCCCATCCCGACGGGCGTAAAGACCCGCCAAAG---AGACAGTAGTAGTCA 1330
 Qy 1255 TCTGACAGTGGAGTTCAAGTAGAGAGCATGGTGAATGCACACAGAGATTCACAGCGG 1314

Db 80 TGCNTTNTATGTTGCTGTGTAGACATGAATAATATCTTCATGATATA 23

RESULT 14
AAK92127
ID AAK92127 standard: cDNA; 807 BP.

XX
XX AAK92127:
AC
XX
XX 06-NOV-2001 (first entry)
DT
XX
XX Human cDNA 5'-end sequence, SEQ ID NO: 587.
DE
XX
XX Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
KW
XX Homo sapiens.
OS
XX
XX EP130094-A2.
PM
XX
XX 05-SEP-2001.
PD
XX
XX 07-JUL-2000; 2000EP-0114089.
PF
XX
XX 08-JUL-1999; 99JP-0194486.
PR
XX 11-JAN-2000; 2000JP-0118774.
PR
XX 02-MAY-2000; 2000JP-0183765.
XX
XX (HELI-) HELIX RES INST.
PA
XX
XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakematsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
XX WPI; 2001-524255/58.
DR
XX
XX 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
PI
XX
XX Claim 2; SEQ ID NO 587; 1380bp + sequence listing; English.
PS
XX
XX The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is the nucleotide
CC sequence of the 5'-end of a cDNA provided in the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
XX Sequence 807 BP; 315 A; 156 C; 163 G; 165 T; 8 other;

Query Match 19.7%; Score 326.4; DB 22; Length 807;
Best Local Similarity 90.3%; Pred. NO. 2.1e-73;
Matches 381; Conservative 0; Mismatches 32; Indels 9; Gaps 3;

QY 2 TGAATTAAGAATATGATACATACACAGAGAAATATCTACAAATGCGCTGAGATGTCAA 61
Db 374 TGAATTAAGAATATGATACATACACAGAGAAATATCTACAAATGCGCTGAGATGTCAA 433
QY 62 TTTATCTAAGTCAACTGGGATTAAGGGTTGAGAGATGAGATGATGCTATCAGCAAAAC 121
Db 434 TTTATCTAAGTCAACTGGGATTAAGGGTTGAGAGATGAGATGATGCTATCAGCAAAAC 493
QY 122 TACATGACCAAGAGAAATATGCGCAAGCTCTATCAGAAATATACATGCAACATATATGCG 181
Db 494 TACATGACCAAGAGAAATATGCGCAAGCTCTATCAGAAATATACATGCAACATATATGCG 553
QY 182 GG--CCAGTACTGGGATTAAGTCTGGGGGAGAGAAACAAAGAGAAACACCTAGAGAA 239
Db 554 GGCCCAAGTACTGGGATTAAGTCTGGGGGAGAGAAACAAAGAGAAACACCTAGAGAA 613

QY 240 TGTCTTAACATATCCGACAGATGATGATTAATGCTAAAGCAGACTCGAA--GGATATA 297
Db 614 TGTCTTAACATATCCGACAGATGATGATTAATGCTAAAGCAGACTCGAAAGGATATA 673
QY 298 AAGAGCCCTCAAGAGATTCGCCAAGCCGAGAAAGTCC-----AGTAAAGCAAAAGCA 352
Db 674 AAGAGCCCTCAAGAGATTCGCCAAGCCGAGAAAGTCCAAATTAAGCAAAAGCA 733
QY 353 CCCATCGTATTCACACACACTGACTAATCAATCTCTCAAAAGTCAAAAGCAAAATCG 412
Db 734 CCCATCGTATTCACACACACTGACTAATCAATCTCTCCCAAAAGTCCAAAGCAAAATCG 793
QY 413 CC 414
Db 794 CC 795

RESULT 15
AAK93391
ID AAK93391 standard: cDNA; 807 BP.

XX
XX AAK93391:
AC
XX
XX 06-NOV-2001 (first entry)
DT
XX
XX Human cDNA clone representative sequence, SEQ ID NO: 1851.
DE
XX
XX Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
KW
XX Homo sapiens.
OS
XX
XX EP130094-A2.
PM
XX
XX 05-SEP-2001.
PD
XX
XX 07-JUL-2000; 2000EP-0114089.
PF
XX
XX 08-JUL-1999; 99JP-0194486.
PR
XX 11-JAN-2000; 2000JP-0118774.
PR
XX 02-MAY-2000; 2000JP-0183765.
XX
XX (HELI-) HELIX RES INST.
PA
XX
XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakematsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
XX WPI; 2001-524255/58.
DR
XX
XX 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
PI
XX
XX Example 11; SEQ ID NO 1851; 1380bp + sequence listing; English.
PS
XX
XX The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence was used as the
CC representative sequence from a human clone which was used in
CC homology searches to identify the clone.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
XX Sequence 807 BP; 315 A; 156 C; 163 G; 165 T; 8 other;

Query Match 19.7%; Score 326.4; DB 22; Length 807;
Best Local Similarity 90.3%; Pred. NO. 2.1e-73;
Matches 381; Conservative 0; Mismatches 32; Indels 9; Gaps 3;

OY	2	TGATTAAGATTAAGTATCGTAAACAAAGAAATACTACAAATGGCCTGAGATGTCAA	61
Db	374	TGATTAAGATTAAGTATCGTAAACAAAGAAATACTACAAATGGCCTGAGATGTCAA	433
OY	62	TTTATTCCTAAGTCAACTGGGGAATTAAGGGTTTGAGATGGAGATGATGCTATCAGCAAAAC	121
Db	434	TTTATTCCTAAGTCAACTGGGGAATTAAGGGTTTGAGATGGAGATGATGCTATCAGCAAAAC	493
OY	122	TACATGACCAAGAGAAATATGGGCGACCTCTCATCAGAAATTAACATATGCAATATATATG	181
Db	494	TACATGACCAAGAGAAATATGGGCGACCTCTCATCAGAAATTAACATATGCAATATATATG	553
OY	182	GG--CCAGTGCATCGATTTAAATCTCTGGGGGAAAGAAACAAAGAGAACACACTTAGGAA	239
Db	554	GGGGCAAGTGCATCGATTTAAATCTCTGGGGGAAAGAAACAAAGAGAACACACTTAGGAA	613
OY	240	TGTTCTAAACATTAATCCGACGACGATATGATTTATGCTTAAAGCACACTCGAA--GGATATA	297
Db	614	TGTTCTAAACATTAATCCGACGACGATATGATTTATGCTTAAAGCACACCTCMAAAGGGATATA	673
OY	298	AAGAAGCCTCAAGAAGATTTCCCAAGCCGACGAAAGGTCC----AGTAAAGGCAAAAGCA	352
Db	674	AAGAAGCCTCAAGAAGATTTCCCAAGCCGACGAAAGGTCCCAATTAAGAAACCAAAAAACAC	733
OY	353	CCCATCGTATTCACACACAACTTGACTTACCTAAAAACATCTCTGCAAAAAGTCAAAAAATCC	412
Db	734	CCCATCGTATTCACACACAACTTGACTTACCTAAAAACATCTCTGCAAAAAGTCCAAAAAATC	793
OY	413	CC 414	
Db	794	CC 795	

Search completed: April 16, 2003, 17:38:00
Job time : 439 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 16, 2003, 10:58:35 ; Search time 2480 Seconds

(without alignments)
10807.875 Million cell updates/sec

Title: US-09-700-696b-1

Perfect score: 1655
Sequence: 1 ggcgaataagaatataagatc.....ataaaaaaaaaaaaaaaaaa 1655

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estdb:*
2: em_esthum:*
3: em_estln:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_estc:*
9: gb_est1:*
10: gb_est2:*
11: gb_est3:*
12: gb_est4:*
13: gb_est5:*
14: gb_est6:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_tod:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	DB	ID	Description
1	219	13.2	571	17	AZ647898	1M0514108
2	154.6	9.3	716	14	BO783292	UI-R-FFO-
3	135.6	8.2	443	10	BB849948	BB849948
4	97.2	5.9	594	14	BO780656	UI-R-FFO-
5	97	5.9	502	10	BE107579	UI-R-CAO-
6	93.4	5.6	499	14	BO782396	BO782396 UI-R-FFO-

7	54.4	3.3	1201	17	CNS0161H	AL106175 Drosophila
8	53.2	3.2	1043	17	CNS0145P	AL103733 Drosophila
9	52.4	3.2	997	17	CNS005TE	AL060767 Drosophila
10	52	3.1	928	17	CNS000DY	AL071865 Drosophila
11	51.6	3.1	1125	9	AL547503	AL547503 AL547503
12	51.6	3.1	1204	17	CNS016E2	AL106628 Drosophila
13	51.4	3.1	568	17	CNS010PL	AL099267 Drosophila
14	51	3.1	1101	17	CNS00XYZS	AL097042 Drosophila
15	50.4	3.0	1059	17	CNS02KFO	AL106415 Drosophila
16	50.2	3.0	1101	17	CNS016DT	AL107172 Drosophila
17	50.2	3.0	1101	17	CNS016T6	AL106619 Drosophila
18	50	3.0	942	17	CNS018GS	AL109318 Drosophila
19	49.8	3.0	404	9	AU061074	AU061074 AU061074
20	49.6	3.0	1101	17	CNS000B8	AL063632 Drosophila
21	49.6	3.0	1101	17	CNS017E8	AL107930 Drosophila
22	49.4	3.0	935	12	BE273928	BE273928 GA-EB001
23	49.4	3.0	988	17	AZ688464	AZ688464 ENTIC86TR
24	49.4	3.0	1001	17	CNS0155H	AL105023 Drosophila
25	49.2	3.0	839	17	CNS0122R	AL101037 Drosophila
26	49.2	3.0	1210	13	B1870038	B1870038 603393904
27	49.2	3.0	839	12	BC856900	BC856900 102404860
28	48.8	2.9	959	17	CNS00655	BC856900 Drosophila
29	48.8	2.9	1124	14	BM545225	BM545225 AGENCOURT
30	48.8	2.9	1184	13	CNS0070G	BM545225 AGENCOURT
31	48.6	2.9	1001	17	CNS0161G	AL066256 Drosophila
32	48.4	2.9	397	17	CNS0161G	AL106174 Drosophila
33	48.4	2.9	806	13	BJ382389	BJ382389 BU382389
34	48.4	2.9	996	17	CNS06HDU	AL398872 T7 end of
35	48.2	2.9	258	12	BC736186	BC736186 FK52g08.Y
36	48.2	2.9	940	17	CNS02GSH	AL196730 Tetradon
37	48.2	2.9	1200	17	CNS016CO	AL106578 Drosophila
38	48	2.9	730	17	AQ879158	AQ879158 HS.3184.B
39	48	2.9	750	17	CNS0111D	AL100303 Drosophila
40	48	2.9	792	17	AZ195555	AZ195555 SP_1030.A
41	48	2.9	890	9	AL581799	AL581799 AL581799
42	48	2.9	1101	17	CNS008X3	AL052544 Drosophila
43	48	2.9	1101	17	CNS00FGC	AL071206 Drosophila
44	48	2.9	1101	17	CNS00GNF	AL072307 Drosophila
45	47.8	2.9	541	17	CNS04KDS	AL295309 Tetradon

ALIGNMENTS

RESULT 1
AZ647898 571 bp DNA linear GSS 14-DEC-2000
1M0514108R Mouse 10kb plasmid UGCGIM library Mus musculus genomic
clone UGCGIM0514108 R, DNA sequence.

ACCESSION AZ647898
VERSION AZ647898.1 GI:11779824
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beecorn,T., Duval,B., Hamill,C.,
Islam,H., Longacire,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D.,Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts

TITLE
JOURNAL
COMMENT Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00

Db	595	ACCTTGACTGAAAGCAAAAGGTTCCACAGCAAAAGGCAAGGCAAG	-----TCT	548
Qy	1033	TCTGTGGTCTTGATTAATGAANTCAAAAACGAATGATTCCTTTAATGCCCCAGTGTAT	1092	
Db	547	TCTCACACTCTTGGTAAAGAGGTTAAAGGTAAAGAAAGACTCTTCAATTAATGTCGTGAGA	488	
Qy	1093	GAGAAATTAATTAACACATGAGCAGAAAATATGATTAATGACCCACAGACAAATTAATCT	1152	
Db	487	GAGGGGATTTGCAAT	-----GCACACAGAGAAACAAGCCACCT	449
Qy	1153	ACACGAAATTAAGGCTATGCCACAAAGGAAAGGCTCTGG--GGTAAACAAACCCATTC	1209	
Db	448	ACACGGAATTAAGGGGATGTACACAGCGAGAGAGCTCTGGCCCTGAGAAAGACCCATGCC	389	
Qy	1210	AACAGAGGTTAGTTCCTCCGTAGAAAGGATGACAGTAAAGTATGATCTGACAGTGGAGT	1269	
Db	388	CACGGCCCGTAAAGACCCGCCCAAG--AGACAGTAACTGATCATCAAGTGGAGT	332	
Qy	1270	TCAAGTGAAGACGATGATGATGACAGTACGACCAAGGATTCACAGCGGGGGTACAGTCTGAAG	1329	
Db	331	TCTAACGAGACAGTGGTGACTAGACACCCCGGGGTTGAATCAAGT	-----	288
Qy	1330	ACCTGCTACCTGTGAGTTGATGTAGAGAGAGACCACCTGACAGCTGACAGGTGAAG	1389	
Db	287	-----	-----CCAGCTCTGCTCTGAGAAAGA	264
Qy	1390	AGGATAGATGAAGACACTGAGTGAAGCCCAAGATCTCTGCTCTCTTGGGGGAATTTTGGT	1449	
Db	263	GAGGACACAGCAGGAGACCTGAGTAAAGTACCAAGACTGGTCAACC-TCCAGGACACCTGGCT	205	
Qy	1450	ATCTTAATAGTCACAGTAAATAATCTTAATTAAGCTTAATGTTTAAGCAAAAAA	1509	
Db	204	GTTTAGTGGTGTAAATAGAAATCCCTCACTCAAAATCTTAAGCTTCTGAATAAAT	145	
Qy	1510	ATCATTAAGATCTTAATGAATAGTAAATGATTAAGTATGATGATTAATTAATAGTTGGT	1569	
Db	144	TTCATTAAGA-ATTATTAATTAATAGTAATATTTGACTAGTGGCCATTAATTAATCTAT	86	
Qy	1570	GAATGTCACAAAATGCCCTTC-TATGTTGTTGCTCTGTAGACATGAATAAACAATATCT	1628	
Db	85	GGATGTCAAAAGTCCCTGATATGATGATTTGCTCTTACAGCATGAATAAAGAGGCTTT	26	
Qy	1629	CTCGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1553	
Db	25	CTTATGAAAAA	-----	1
RESULT 3				
LOCUS	BB849948			
DEFINITION	BB849948 RIKEN full-length enriched, adult inner ear Mus musculus	443 bp	linear	EST 26-NOV-2001
ACCESSION	BB849948			
VERSION	BB849948			
KEYWORDS	BB849948.1	GI:17091402		
SOURCE				
ORGANISM		house mouse.		
REFERENCE				
AUTHORS				
TITLE				
COMMENT				

[illegible]

RESULT 4

B0780656/c

LOCUS B0780656 594 bp mRNA linear EST 26-JUL-2002

DEFINITION UI-R-F00-cpc-c-14-0-UI.s1 UI-R-F00 Rattus norvegicus cDNA clone

ACCESSION B0780656

VERSION B0780656.1 GI:21989128

KEYWORDS EST..

SOURCE Norway rat.

ORGANISM

Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 594)

AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene

JOURNAL

MEDLINE Genome Res. 6 (9), 791-806 (1996)

COMMENT

Contact: Soares, MB

Program for Rat Gene Discovery and Mapping

University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

Tissue Procurement: Jeff Stevens

cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa

cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: DISTRIBUTION: Researchers may obtain clones

from Research Genetics (www.resgen.com).

Seq primer: M13 FORWARD

POLYA-Yes.

Location/Qualifiers

FEATURES

source

1..594

/organism="Rattus norvegicus"

/db_xref="taxon:10116"

/clone="UI-R-F00-cpc-c-14-0-UI"

/clone_lib="UI-R-F00"

/tissue_type="Mixed tissues"

/dev_stage="Adult"

/lab_host="DH10B (Life Technologies) (T1 phage resistant)"

/note="Vector: pUT3-Pac (Pharmacia) with a modified

polylinker. Site_1: EcoR I. Site_2: Not I; UI-R-F00 is a

subtracted cDNA library containing the following tissue(s)

): Normal cartilage and SR-JWS Tumor Line. The

subtraction was made according to Bonaldo, Lennon and

Soares, Genome Research, 6:791-806, 1996. The

oligonucleotide used to prime the synthesis of

first-strand cDNA contains a library tag sequence that is

located between the Not I site and the (dT)18 tail. The

sequence tags for these libraries are: CTAATGACG,

CAATCTGTGA.

TAG_LIB-UI-R-F00

TAG_TISSUE=cartilage

TAG_SEQ=CTAATGACG"

BASE COUNT 132 a 149 c 123 g 190 t

ORIGIN

Query Match

Best Local Similarity 57.5%; Score 97.2; DB 14; Length 594;

Matches 359; Conservative 0; Mismatches 183; Indels 82; Gaps 7;

OY 1031 CTCTCGTGTCTGTGATGAATCAAAAGCAAGATGCTTTAATGCCCCCACTC 1090

DB 546 CTCTCAAGCTCTGTGATGAATGAAGTAAAGAAAGACTCTCTAATAGTCTCACTA 487

OY 1091 ATGGAATATATATACACATGCGAGAAATATCTATATGTAACCCACAGCAAAATATT 1150

DB 486 GAGAGGGGATTGCAATA-----GCACACAGAGGAGCAAGCCACC 448

OY 1151 CTACACGATATAGGCTATGCCACAGCAAAAGCTCTGGGGT---AGACACCCCAT 1207

DB 447 CTACACGATATAGGCGATGTCACAGCGAGAGGCTCTGGCCCTCGAGAAAGACCCCATC 388

OY 1208 CCACAGAGGCTTTAGTCCCGTAGAAGGATGACAGTAGTAGTCATCTGACAGTGCGA 1267

DB 387 CCCACCGCGCGGTAGACACCCGCCAAGA---GACAGTAGTAGTCATCTGACAGTGCGA 331

OY 1268 GTTCAAGTAGACGATGCTGCTCCACAGGATTCACCGGGGTGACAGTCTGA 1327

DB 330 GTTCAAGTAGACGATGCTGCTCCACAGGATTCACCGGGGTGACAGTCTGA 285

OY 1328 AGACCTGTCACCTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1387

DB 284 -----CCACGCTCTGCTCTGGAGAG 263

OY 1388 AGAGATGAGTGAAGACGATGAGGACCAAGATCTGCTCTGCTGCTGCTGCTGCTG 1447

DB 262 GAGAGACACAGCAGGAGCTGATGATGATGATGATGATGATGATGATGATGATGATG 204

OY 1448 CTATCTTAATGCTGACATATATAATTTCTAATGAGCTATATGTTTAAAGCAAAA 1507

DB 203 CTGTTTAGTGTGTATATAGAAATCCCTACTCAAGTTCTAATGCTTTCTGAATTA 144

OY 1508 AAATCATTAAGATCTATGAATAGTATGATGATGATGATGATGATGATGATGATGATG 1567

DB 143 TTTTCATAGA-ATTATATATATAGTATATTTAGACAGGTGGCCCTTAATATAGCT 85

OY 1568 GTGAATGCTCAAAATGCTCT-CTATGTTGTTGCTCTGATGATGATGATGATGATG 1626

DB 84 ATGATGCTCAAGAGGCTCTGATGATGATGATGATGATGATGATGATGATGATGATG 25

OY 1627 CTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1650

DB 24 TTCTCTAAAAAATAAAAAA 1

RESULT 5

BE107579/c

LOCUS BE107579 502 bp mRNA linear EST 13-JUN-2000

DEFINITION UI-R-CA0-aw-d-02-0-UI.s1 UI-R-CA0 Rattus norvegicus cDNA clone

ACCESSION BE107579

VERSION BE107579.1 GI:8499684

KEYWORDS EST.

SOURCE Norway rat.

ORGANISM

Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 502)

AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene

JOURNAL

MEDLINE Genome Res. 6 (9), 791-806 (1996)

COMMENT

Contact: Soares, MB

Program for Rat Gene Discovery and Mapping

University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

The sequence contained an oligo-dT track that was present in the

oligonucleotide that was used to prime the synthesis of first

strand cDNA and therefore this may represent a bonafide poly A

tail. cDNA library Preparation: M.B. Soares Lab Clone distribution:

clones will be available through Research Genetics (www.resgen.com)

Seq primer: M13 Forward

POLYA-Yes.

Location/Qualifiers

FEATURES

source

1..502

```

/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-CA0-aw-d-02-0-UI"
/lab_host="UI-R-CA0"
/lab_host="DH10B (Life Technologies)"
/Note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; The UI-R-CA0
library is a subtracted library derived from the following
tissues: thalamus, cerebellum, hypothalamus, medulla, pons
, midbrain, cerebral cortex, corpus striatum, testis, and
hippocampus. For a detailed description of the library
from this clone was derived, please visit our web
site at atasest.eng.uiowa.edu. The subtraction has been
previously described in (Bonaldo, Lennon and Soares,
Genome Research 6:791-806, 1996)
TAG_SEQ=None found"
BASE COUNT      112 a      128 c      104 g      157 t      1 others
ORIGIN

```

```

Query Match      5.9%; Score 97; DB 10; Length 502;
Best Local Similarity 58.9%; Pred. No. 3.5e-08;
Matches 311; Conservative 0; Mismatches 156; Indels 61; Gaps 6;

```

```

QY 1129 GATCCACACAGCAAAATATTTCTACACGGAATAGGGATGGCCACAGGAAAGGCTCC 1188
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 471 GCACACAGAGAACACACACACACACACACACACACACACACACACACACACACACAC 412
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1189 TGGG---GTAGACAAACCCCTTCCACAGAGGTTTACGTTCCCGTAAAGGATGACAGT 1245
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 411 TGGGCTCGAGAGACCCCTCCACCGGCGGTAGACCCCGCCCAAGCA---GACAGT 355
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1246 AGTAGTCATCTGACAGTGGCAGTCAAGTGAAGAGGATGATGATGATGATGATGATGAT 1305
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 354 AGTAGTCATCTGACAGTGGCAGTCTAGCAGAGACAGAGTGTACTAGACACCCGGGCTTG 295
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1306 TCCACGCGGGGTGACAGTGTGAAGACCTGCTGATGATGATGATGATGATGATGATGAT 1365
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 294 AACCACTT----- 287
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1366 CCTGACAGCTGACAGGATGAAGAGATAGATGAAGAACTAGTGAAGCAAGATCCT 1425
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 286 CCCAGCTCTGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 227
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1426 GGTCTCTGGGGAAATTTTCTATCTTAATAGTACAGTATATAAATTTATTAAGGC 1485
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 226 GGTCAAC-TCACAGACACGTGCTGTTTAACTGTTTAACTGTTTAACTGTTTAACTGTT 168
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1486 TATATGTTTAAAGCAAAATATCTATGATGATGATGATGATGATGATGATGATGATGAT 1545
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 167 TCTAATG-CTTCTGTAATTAATCTTCTGTAATTAATTAATTAATTAATTAATTAATTA 109
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1546 AGGTGATCTTAAATATGTTGTAATGTCACAAATGCTT-CTATGTTGTTTGTCTG 1604
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 108 AGGCGCCATTAATTAATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 49
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1605 TAGACATGAATAATTAACATATCTCTGATGATTAATAAATAAATAAATAAATAA 1652
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 48 CAGACATGAATAAAGAGGCTTCTCTATGAATAAATAAATAAATAAATAAATAAATAA 1
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 6
B0782396/c      499 bp      mRNA      linear      EST 26-JUL-2002
LOCUS           B0782396
DEFINITION     UI-R-F00-cpk-e-05-0-UI.s1 UI-R-F00 Rattus norvegicus cDNA clone
ACCESSION      B0782396
VERSION        B0782396.1 GI:21990868
KEYWORDS
SOURCE         Norway rat.
ORGANISM       Rattus norvegicus
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

```

```

REFERENCE
1 (bases 1 to 499)
AUTHORS        Ronaldo,M.F., Lennon,G. and Soares,M.B.
TITLE          Normalization and subtraction: two approaches to facilitate gene
               discovery
JOURNAL        Genome Res. 6 (9), 791-806 (1996)
MEDLINE        97044477
COMMENT        Contact: Soares, MB
               Program for Rat Gene Discovery and Mapping
               University of Iowa
               451 Eckstein Medical Research Building Iowa City, IA 52242, USA
               Tel: 319 335 8250
               Fax: 319 335 9565
               Email: msoares@blue.weeg.uiowa.edu

```

```

Tissue Procurement: Jeff Stevens
CDNA library preparation: Dr. M. Bento Soares, University of Iowa
CDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: DISTRIBUTION: Researchers may obtain clones
from Research Genetics (www.resgen.com).
Seq primer: M13 FORWARD
POLYA-Tes.

```

```

FEATURES
source
Location/Qualifiers
1..499

```

```

/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="UI-R-F00-cpk-e-05-0-UI"
/lab_host="UI-R-F00"
/tissue_type="Mixed tissues"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/Note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site_1: Eco RI; Site_2: Not I; UI-R-F00 is a
subtracted cDNA library containing the following tissue(s)
: Normal cartilage and SR-JMS Tumor line. The
subtraction was made according to Ronaldo, Lennon and
Soares, Genome Research, 6:791-806, 1996. The
oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tags for these libraries are: CTATGAGACG,
CATCTCTGTA,
TAG_LIB=UI-R-F00
TAG_TISSUE=cartilage
TAG_SEQ=CTATGAGACG"
BASE COUNT      112 a      123 c      103 g      160 t      1 others
ORIGIN

```

```

Query Match      5.6%; Score 93.4; DB 14; Length 499;
Best Local Similarity 58.6%; Pred. No. 1.6e-07;
Matches 308; Conservative 0; Mismatches 157; Indels 61; Gaps 6;

```

```

QY 1129 GATCCACACAGCAAAATATTTCTACACGGAATAGGGATGGCCACAGGAAAGGCTCC 1188
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 469 GCACACAGAGAACACACACACACACACACACACACACACACACACACACACACACAC 410
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1189 TGGG---AGACACCCCATTCACACAGAGGTTTACGTTCCCGTAAAGGATGACAGT 1245
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 409 TGGGCTCGAGAGACCCCATTCACACAGAGGTTTACGTTCCCGTAAAGGATGACAGT 353
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1246 AGTAGTCATCTGACAGTGGCAGTCAAGTGAAGAGGATGATGATGATGATGATGATGAT 1305
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 352 AGTAGTCATCTGACAGTGGCAGTCTAGCAGAGACAGTGTACTAGACACCCGGGCTTG 293
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1306 TCCACGCGGGGTGACAGTGTGAAGACCTGCTGATGATGATGATGATGATGATGATGAT 1365
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 292 AATCAGTT----- 285
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1366 CCTGACAGCTGACAGGATGAAGAGATAGATGAAGAACTAGTGAAGCAAGATCCT 1425
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 284 CCCAGCTCTGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 225
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1426 GGTCTCTGGGGAAATTTTCTATCTTAATAGTACAGTATTAATAATTTATTAAGGC 1485
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db	224	GGTCACC-TCGAGCAGCATGCTGCTTTAGTGGTGTGTAATAGATCCCTACCTCAAGT	166
Qy	1486	TATATATGTTTTAAAGCAAAAAAAATCATTTACAGATCATGAATAGGTACATTTGAGT	1545
Db	165	TCTAATGCTTTCTGATTAATAAATTTCTCAAGA-ATTATATTAATAGGTAATATTGNNCT	107
Qy	1546	AGGCTCATTTAAAAATAGTGGTGAAGTCACAAATGCCCT-CTATGTTGTTGCTCTCG	1604
Db	106	AGGTGGCCCATTAATATGTCTATGAGATGCACAAGTCCCTGATATGTGATGCTCTT	47
Qy	1605	TAGCATGAATAATTAACAATATCTGCATATATAAAAAAAAAAAAAA	1650
Db	46	CAGCATGAAAAATTAAGAGGCTTCTCTTAATAAAAAAAAAAAAAAAAAA	1

RESULT 7	
CNS0161H/C	1201 bp DNA linear GSS 26-JUL-1999
LOCUS	
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC
	BAH15018 of DrosBC library from Drosophila melanogaster (fruit
	fly), genomic survey sequence.
ACCESSION	AL106175
VERSION	AL106175.1 GI:5620521
KEYWORDS	GSS.
SOURCE	Drosophila melanogaster.
ORGANISM	Drosophila melanogaster.

REFERENCE
AUTHORS
TITLE
JOURNAL

Ephydroidae: Drosophilidae: Drosophila.
1 (bases 1 to 1201)
Genoscope.
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : :

COMMENT
Determination of this BAC-end sequence was carried out as part of collaboration with the European Drosophila Genome Project (EDGP) - <http://www.edgp.ebi.ac.uk/>. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billard at CESP (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.

FEATURES	SOURCE	Location/Qualifiers
	1.	1201
		/organism="Drosophila melanogaster"
		/db_xref="taxon:7227"
		/clone="BACN15018"
		/clone_1ib="DrosBAC"
		/plasmid="pBelobAC11"
		/note="end : 77"
BASE COUNT	120 a	163 c 157 g 513 t 248 others
ORIGIN		

	Query Match	Similarity	3.3%	Score	54.4	DB	17	Length	1201
	Best	Local	Similarity	37.2%	Pred.	No.	1.9		
	Matches	151	Conservative	47	Mismatches	208	Indels	0	Gaps
QY	4	AATAAGATATATAGTTCGTCTGCTAACAAAGAGATATCTCCAAATGGCGCTGGAGATGCAATT	63						
Db	504	AA	445						
OY	64	TATCTTAAGTCAACTCGGAGATTAAGGGTTGGAGATGGAGATGATGCTATCGACCAACTA	123						
Db	444	AA	385						
OY	124	CATGACCAAGAAATATGTGGCCAGCTCTCATCTCAGAAATAATACATGCAACATATAATGGGG	183						
Db	384	AA	325						
OY	184	CCAGTCACTGCATTAAACTCTCTGGGGGAGAGAAACAAAGAGACACACCTAGAAATGTT	243						

Dd	324	AATARAATAAACAAAMMAWATGTOMAAAAATAMAAAAAAMMMMAAAAAMAAACAAWWAAAWAA	265
Qy	244	CTAAACATATATCCAGCAGAATGATGAATTATGCTTAAGCACACTCGAAGAGTAAAAAGAG	303
Dd	264	RAAAMMAAAWACMTAAAMMAAATTAATAAAMMMAAMCAAMMAATTAAMMAAAAAAAAAWTGRAA	205
Qy	304	CCTCAAAGAGTTCCCAAGGCCACAAAAAGTCACATTAATAAAAAACAACACCACCTGATT	363
Dd	204	AAWTAAAMARMAAMAMMMRAACMAAATTRGRMAAAAAAAAAAAAAAAAAAAAAAAAAAMM	145
Qy	364	CAACGACACATTGCATCTAAACATCTCTCAAAAGTCAAAAA	409
Dd	144	AMAAAAAAAAAATACACATATATTAAGMMAAAAAAAAAAAAAAAAAMMMMMAAAWA	99

RESULT 8	
CNS0145P/C	
LOCUS	1043 bp DNA linear GSS 26-JUL-1999
DEFINITION	CNS0145P
	Drosophila melanogaster genome survey sequence T7 end of BAC
	BACN11G1 of DrosBAC library from Drosophila melanogaster (fruit
	fly) genomic survey sequence.
ACCESSION	AF103735
VERSION	GI:5615346
KEYWORDS	GSS.
SOURCE	Drosophila melanogaster.
ORGANISM	Drosophila melanogaster.

REFERENCE	AUTHORS	TITLE	JOURNAL
1 (bases 1 to 1043)	Genoscope.	Direct Submission	Submitted (23-JUL-1999)
Genoscope - Centre National de Sequencage : http://www.genoscope.cnr.fr			

COMMENT
Denaturation of this BAC-end sequence was carried out as part of collaboration with the European Drosophila Genome project (EDGP) - <http://www.edgp.ebi.ac.uk/>. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billard at CGEP (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.

FEATURES	source	Location/Qualifiers
		1..1043
		/organism="Drosophila melanogaster"
		/db_xref="taxon:7227"
		/clone="BACN1IG11"
		/clone_lib="DrosBAC"
		/plasmid="pbeloBAC11"
		/note="end : 17"
BASE COUNT	277 a	96 c 121 g 382 t 167 others
ORIGIN		

[illegible]

RESULT 9
 CNS005TE/C
 LOCUS
 DEFINITION CNS005TE 997 bp DNA linear GSS 03-JUN-1999
 Drosophila melanogaster genome survey sequence TE73 end of BAC #
 BACR12K22 of RPCI-98 library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.
 ACCESSION
 VERSION AL060767.1 GI:4943573
 KEYWORDS
 SOURCE GSS.
 ORGANISM Drosophila melanogaster.
 Drosophila melanogaster.
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 997)
 REFERENCE
 AUTHORS Direct Submission
 TITLE Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
 JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)
 COMMENT Determination of this BAC end sequence was carried out as part of a
 collaboration with the Berkeley Drosophila Genome Project (BDGP).
 The BDGP is constructing a physical map of the Drosophila
 melanogaster genome using these BACs. For further information
 please see <http://www.fruitfly.org> The BDGP Drosophila
 melanogaster BAC library was prepared by Kazutoyo Osoegawa and
 Aaron Mamoser in Pieter de Jong's laboratory in the Department of
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
 NY. The library is named RPCI-98 and was constructed by partial
 EcoRI digestion of Drosophila DNA provided by the BDGP from the
 isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
 p1 and EST libraries. A more detailed description of the library
 and how to order individual BAC clones, the entire library, or
 filters for hybridization from the BACPAC Resource Center can be
 found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
 FEATURES
 source
 1..997
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone="BACR12K22"
 /clone_1lb="RPCI-98"
 /note="end : TE73"
 BASE COUNT 89 a 99 c 13 g 258 t 538 others
 ORIGIN
 Query Match 3.2%; Score 52.4; DB 17; Length 997;
 Best Local Similarity 16.6%; Pred. No. 4.7;
 Matches 62; Conservative 137; Mismatches 175; Indels 0; Gaps 0;

Db 509 GARRRGRRGRRRAAGARRRAAARARRARAGARRRRRTNNNNNNNNNN 450
 QY 964 AACCAAGCAACCTT 977
 Db 449 NNNNNAGCNTNNNT 436
 RESULT 10
 CNS00DKY/C
 LOCUS
 DEFINITION CNS00DKY 928 bp DNA linear GSS 04-JUN-1999
 Drosophila melanogaster genome survey sequence T7 end of BAC #
 BACR27A24 of RPCI-98 library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.
 ACCESSION
 VERSION AL071865.1 GI:4948170
 KEYWORDS
 SOURCE GSS.
 ORGANISM Drosophila melanogaster.
 Drosophila melanogaster.
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 928)
 REFERENCE
 AUTHORS Direct Submission
 TITLE Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
 JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)
 COMMENT Determination of this BAC end sequence was carried out as part of a
 collaboration with the Berkeley Drosophila Genome Project (BDGP).
 The BDGP is constructing a physical map of the Drosophila
 melanogaster genome using these BACs. For further information
 please see <http://www.fruitfly.org> The BDGP Drosophila
 melanogaster BAC library was prepared by Kazutoyo Osoegawa and
 Aaron Mamoser in Pieter de Jong's laboratory in the Department of
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
 NY. The library is named RPCI-98 and was constructed by partial
 EcoRI digestion of Drosophila DNA provided by the BDGP from the
 isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
 p1 and EST libraries. A more detailed description of the library
 and how to order individual BAC clones, the entire library, or
 filters for hybridization from the BACPAC Resource Center can be
 found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
 FEATURES
 source
 1..928
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone="BACR27A24"
 /clone_1lb="RPCI-98"
 /note="end : T7"
 BASE COUNT 262 a 70 c 84 g 321 t 191 others
 ORIGIN
 Query Match 3.1%; Score 52; DB 17; Length 928;
 Best Local Similarity 35.6%; Pred. No. 5.7;
 Matches 77; Conservative 47; Mismatches 92; Indels 0; Gaps 0;

REFERENCE
AUTHORS
TITLE
JOURNAL

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1204)

COMMENT

Genoscope.
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
library (www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC
library (DrosBAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
peloBAC11.

FEATURES
source

Location/Qualifiers
1. .1204
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACN15A12"
/clone_1lb="DrosBAC"
/plasmid="pBelOBAC11"
/note="end : T7"

BASE COUNT 298 a 172 c 106 g 316 t 312 others

ORIGIN

Query Match 3.1%; Score 51.6; DB 17; Length 1204;
Best Local Similarity 22.4%; Pred. No. 6.3;
Matches 70; Conservative 118; Mismatches 124; Indels 0; Gaps 0;

QY 1341 TGTGAGTTGATGTAGAGAGAGACCACTGACAGCTGACCAGTGAAGAGAGATAGAGTG 1400
Db 1055 KKAKADAKKADKKDKDKAKAKADAMDDADAKKKKKAKKKKKDDATDADAAAKAKK 996
QY 1401 AAGAAGCTAGAGAGCCAGCAAGATCCGCTGCTCGTGGGGAATTTTGCATCTTAATAGT 1460
Db 995 DKAAKKKKKKDKKAKKAAKAAKKKKDKKKDKKKKKKKKKKKKKKAAKDAAAAAA 936
QY 1461 CACAGATATAAAATTTCTATTAAGGCTATATGTTTAAAGCAAAAAAATCATTAACGA 1520
Db 935 KKKDKKDAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 876
QY 1521 TCTATGAATATGCTAAGACATTGTAGAGAGTGTCATTTAAATAAGTTGGTGAATGCACAA 1580
Db 875 KKAADAKAAKKKKDKKAAKADDAAKKKKKKAAAGAAARAAAKGGGGRKGAGGGGA 816
QY 1581 ATGCCTTCTATGTGTGTTGGCTGTAGACATGAATAAATATCTCTCGATGATATAA 1640
Db 815 AAAGATTTTAAKTDITTKKAGDAAKTTTTTKTTTAAKAGAAKAKKTTRAAAKDARCKDAAT 756
QY 1641 AAAAAAAAAA 1652
Db 755 TAKAARAAGAAA 744

RESULT 13
CNS010PL/C
LOCUS
DEFINITION

CNS010PL 568 bp DNA linear GSS 26-JUL-1999
Drosophila melanogaster genome survey sequence SP6 end of BAC
BACN04L10 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
AL099267
AL099267.1 GI:5610878
GSS.

ACCESSION
VERSION
KEYWORDS
SOURCE

Drosophila melanogaster.
Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 568)
Genoscope.
Direct Submission

REFERENCE
AUTHORS
TITLE

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1204)

Thu Apr 17 09:34:34.2003

us-09-700-696b-1.rst

```

/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"

```

```

/ab_xiel= .laou.3200
/c1one="145K18"

```

```
/clone_lib="G"
```

```

/c10ne_id=9
/note="Genoscope sequence ID : C0AG145BF09SP1-end :
nuc-ori"

```

PUC-OR1™

a	153 c	175 g	170 t	66 others
---	-------	-------	-------	-----------

BASE COUNT	495 a	153 c	175 g	170 t	88 OTHERS
ORIGIN					

Query Match	3.0%;	Score 50.4;	DB 17;	Length 1059;
Best Local Similarity	44.1%;	Pred. No. 11;		

3.08;	Score 50.4;	DB 17;	Length 1059;
44.18;	Pred. No. 11;		

Query Match	3.0%;	DB 17	Length 100;
Best Local Similarity	44.1%;	Pred. No. 11;	
Matches 150; Conservative	22;	Mismatches 163;	Indels 5; Gaps 1;

servative	22;	Mismatches	163;	Indels	5;	Gaps	17
-----------	-----	------------	------	--------	----	------	----

70 AAGTCACTGGGAATAAGGTTGACGATGGAGATGATGCTATCAGCAACTACATGAC 129

Db 190 AAAAAARCMGCACMAAACTRGAWAAGGAARAAAGAAMGAAGAAGAAAAAAAMCRRAA 249

130 CAAGAAGATATGGCCAGCTCTCATCAGAAATACATGCACATATATGCGCCAGT 189

350 A A A A M A G A A R A G A W G A A R A A A A M A A A A G A A R W A A R G A A A M A A W A A A A A A A G A T T T T A M 309

100. TATGCGCTATTAATCCCTGCGCGGAGAGAAACAAAGAGACACACCTTGGATGTTCTAAC 249

[illegible]

D_b 310 ACAGGAGAAAMAGAAGAGAAAAAATTTTCTTCCTTTTTTGGCATTGTCCACCGCATTAATAAGAACCTCAAA 309

250 ATATCCAGCAGTATGATTATGCTAAAGCACACACCTCGAAGGATTTTAAAAAACCAATTA
||| ||| : ||| ||||| |

Db 370 AAAAAMCGAAAAAGARAAAAAAAAAAAAAAAAAAAAAARC-----AAAGAAAAAAAAAGAGCNCMAAAA *2*

310 AGAGATCCCAAGCCAGAAAGTCCAGTAAAAAGCAAAAGCACCATCGTATTCACAC 365

Db 425 AAAAAAAAAAAAAAAAAA.TRAARARAAAAAAMMAAAAAARAGATAAAAAAA 48

370 AACATTGACTACCTAAACACATCTCTCAAAAGTCAAAAAA 409

485 AAAAAAAAAAARAAAAAAAAAAGAAAAAATATATATAAA 524

Search completed: April 16, 2003, 19:35:02
Job time : 2511 secs

Job time : 2511 secs

Thu Apr 17 09:34:34 2003

us-09-700-696b-1.rni

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 16, 2003, 13:05:58 Search time 75 Seconds
(without alignments)
6767.334 Million cell updates/sec

Title: US-09-700-696b-1

Perfect score: 1655
Sequence: 1 gtcgaataagaatagatc.....ataaaaaaaaaaaaaa 1655

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Issued Patents-NA: *
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/1/ina/5A.COMB.seq: *
4: /cgn2_6/ptodata/1/ina/5B.COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PTCUS.COMB.seq: *
6: /cgn2_6/ptodata/1/ina/Deckfile1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	70	4.2	7218	1	Sequence 14, Appl
2	48	2.9	2277	1	Sequence 2, Appl
3	48	2.9	2277	1	Sequence 2, Appl
4	48	2.9	2277	1	Sequence 2, Appl
5	42.2	2.5	2621	2	Sequence 193, App
6	39.6	2.4	696	4	Sequence 187, App
7	39.6	2.4	699	4	Sequence 189, App
8	39.6	2.4	717	4	Sequence 185, App
9	39.6	2.4	714	4	Sequence 184, App
10	39.6	2.4	1669	3	Sequence 11, Appl
11	39.6	2.4	719	4	Sequence 11, Appl
12	39.2	2.4	719	4	Sequence 11, Appl
13	39.2	2.4	719	4	Sequence 11, Appl
14	39.2	2.4	719	4	Sequence 11, Appl
15	39.2	2.4	1450	2	Sequence 32, Appl
16	39.2	2.4	2447	2	Sequence 14, Appl
17	39.2	2.4	1857	2	Sequence 4, Appl
18	39.2	2.4	4098	4	Sequence 132, Appl
19	38	2.3	2773	2	Sequence 169, App
20	38	2.3	2007	2	Sequence 1710, Ap
21	38	2.3	2007	2	Sequence 1710, Ap
22	38	2.3	2028	4	Sequence 169, App
23	38	2.3	2422	1	Sequence 5, Appl
24	38	2.3	3138	1	Sequence 4, Appl
25	38	2.3	5852	1	Sequence 2, Appl
26	37.6	2.3	1293	4	Sequence 219, App
27	37.4	2.3	3527	2	Sequence 7, Appl

28	37.2	2.2	1474	4	US-08-821-994-64	Sequence 64, Appl
29	36.8	2.2	1038	4	US-09-560-419-4	Sequence 4, Appl
30	36.8	2.2	2673	4	US-09-336-447A-12	Sequence 12, Appl
31	36.4	2.2	4285	4	US-09-410-464-1	Sequence 1, Appl
32	36.2	2.2	681	5	PCT-US95-13658-3	Sequence 2725, Ap
33	36.2	2.2	870	5	US-09-134-001C-2725	Sequence 3, Appl
34	36.2	2.2	11288	4	US-08-646-301A-1	Sequence 1, Appl
35	36.2	2.2	11288	4	US-08-481-968A-4	Sequence 4, Appl
36	36.2	2.2	11288	4	US-08-154-712B-4	Sequence 4, Appl
37	36.2	2.2	30549	4	US-09-134-001C-322	Sequence 322, App
38	36	2.2	1219	4	US-08-195-705-1	Sequence 1, Appl
39	36	2.2	3396	2	US-08-974-565C-6	Sequence 6, Appl
40	36	2.2	3396	2	US-09-255-748-6	Sequence 6, Appl
41	36	2.2	5181	1	US-08-257-073-10	Sequence 10, Appl
42	35.8	2.2	188	1	US-08-470-179-44	Sequence 44, Appl
43	35.6	2.2	222	4	US-08-481-190-15	Sequence 15, Appl
44	35.6	2.2	222	5	PCT-US93-00869-15	Sequence 15, Appl
45	35.4	2.1	731	1	US-08-451-405A-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14/C
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEFFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)836-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZgpt-F15
US-08-232-463-14
Query Match 4.2%; Score 70; DB 1; Length 7218;

2.9%: Score 48. Dp 1. Total 0000

ADDRESSEE: Science & Technology Law Group
STREET: 268 Bush Street, Suite 3200
CITY: San Francisco

ADDRESSEE: Science & Technology Law Group
STREET: 268 Bush Street, Suite 3200
CITY: San Francisco

COUNTRY: USA
 ZIP: 94104
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: IBM PC compatible
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/676,974
 FILING DATE:
 CLASSIFICATION: 530
 ATTORNEY/AGENT INFORMATION:
 NAME: Osman Ph.D., Richard A
 REGISTRATION NUMBER: 36,627
 REFERENCE/DOCKET NUMBER: UC96-055
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415)343-4341
 TELEFAX: (415)343-4342
 INFORMATION FOR SEQ. ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2277 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 US-08-676-974-2

Query Match .2.9%; Score 48; DB 1; Length 2277;
 Best Local Similarity 26.1%; Pred. No. 0.0024;
 Matches 205; Conservative 119; Mismatches 458; Indels 3; Gaps 1;

226 AACGACCTGAGATGTTCTTAACATATCCCGACAGTATGATTAATGCTAAAGCAC 285
 214 AARTHAAYGTNACNGTNGCNAARAARATYTNMGNAIYARACNARAGARAGGAAR 273
 286 TCGAAGCATAAAGAGCTCAAGAGATTCCCAAGCCGAGAAAGTCCAGTAAAGC 345
 274 AAYGARAAYNSNGARTGYCCNAARAARGCNAARAGCNAARAARGNAGTNGAY 333
 346 AAAAGCACCATGCTATTCAACACACATGACTACTAAACATCTCTCAAAAGTAAA 405
 334 AARAARCGNMGNTNATHATMGNAAYTTNMSNTTYAARTGYWSNGARAGAYTTAAR 393
 406 AAAATCCCGAGTATTTGAAGCAGCGGTATACAGATCTCTCAAGAGAGGAGCAAT 465
 394 ACNGTNTTTCNCARTTYGGNGCGTNTYTNARGTNAAYATHCNMGNAARCCGAGCN 453
 466 GATATATCTCTTCACTGAGGAGCGCAACCTTTTAAGGACATCTCTGTAAGAGAA 525
 454 AARATGNGNGNT---YGGTNTYGTNCARTTYAARAAYTYNTYNGARCGNGNAARCN 510
 536 GCTACTGCTGACCTTAGAAGCAAGATTTTCAACAGGCTTGCAGGCCCAAGTAA 585
 511 YTNARAGNATGAAYATGAARATTAARAGNMGNCNGTNGTNGATTTGGGCGTIN 570
 586 GCTGAGTACTCATCTTGACACAAAAAGCCAGGTTATATGATGATCCAGAGAGAA 645
 571 GCNARAGYAAARTYAANGATACNCARNSGNTNMSNGCNATHGNGARAGAAAMWSNAY 630
 646 GAAATGCTGGAATACCATTTGAACTAGAGGATGAACTGCGAAGAGCAGATGCTGT 705
 631 GARMSNARCAVCAARGAWSNGTNAARAARAARGNMGNGARAGAGATATGARGAR 690
 706 GATGACACCTTTGAGAGGACAGACATATCATGGTATACCAATTTTAAGAGCTC 765
 691 GARGARAYGAYGAYGAGAGAGAYGAYGAYGAYGAYGAYGAYGAYGAYGAYGAR 750
 766 CTTGGAAGAGAAAGACAGAGTGTGTCAGAGCCAAATGCTCACCAGAGGAAGTT 825
 751 GAYGARAGARAGARAAATATGACARMSNARGTNMCNARNGTNCATATTCARABRGN 810
 826 GAGTTTCATTAACCTCTGACCTCTCAAAAGAGAAAGAGGAGGAGGAGCAAT 885

DB 811 GCGTNAARMGNCNGCNCNCNARWSNWSNGAYCAYWSNGARAGARAGAYWSNGAYTN 870
 QY 886 GCTGAAGTACCACTATAATGAATTCCTTAATGCAAGGACAGTACAGAAAGGT 945
 DB 871 GARGARWSNGAYWSNATGAGAGGNGARARRTYNGCNCARWSNGATACMWSNACNGAR 930
 QY 946 GATGATCTTCTAATAGGACCAACCTTAATGAAAAACAAGTTCTCTAGTAA 1005
 DB 931 GARGARAGAYAAAGCNGTNCARTNMSNAIYARARARARARARAYTTCNMWSNAY 990
 QY 1006 GGCAA 1010
 DB 991 GTNAA 995

RESULT 4
 US-09-098-487-2
 Sequence 2, Application US/09098487
 Patent No. 5917025
 GENERAL INFORMATION:
 APPLICANT: COLLINS, Kathleen
 TITLE OF INVENTION: Human Telomerase
 NUMBER OF SEQUENCES: 11
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Science & Technology Law Group
 STREET: 268 Bush Street, Suite 3200
 CITY: San Francisco
 STATE: CA
 COUNTRY: USA
 ZIP: 94104
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: IBM PC compatible
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/098,487
 FILING DATE:
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Osman Ph.D., Richard A
 REGISTRATION NUMBER: 36,627
 REFERENCE/DOCKET NUMBER: UC96-055
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415)343-4341
 TELEFAX: (415)343-4342
 INFORMATION FOR SEQ. ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2277 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 US-09-098-487-2

Query Match .2.9%; Score 48; DB 2; Length 2277;
 Best Local Similarity 26.1%; Pred. No. 0.0024;
 Matches 205; Conservative 119; Mismatches 458; Indels 3; Gaps 1;

226 AACGACCTGAGATGTTCTTAACATATCCCGACAGTATGATTAATGCTAAAGCAC 285
 214 AARTHAAYGTNACNGTNGCNAARAARATYTNMGNAIYARACNARAGARAGGAAR 273
 286 TCGAAGCATAAAGAGCTCAAGAGATTCCCAAGCCGAGAAAGTCCAGTAAAGC 345
 274 AAYGARAAYNSNGARTGYCCNAARAARGCNAARAGCNAARAARGNAGTNGAY 333
 346 AAAAGCACCATGCTATTCAACACACATGACTACTAAACATCTCTCAAAAGTCAA 405
 334 AARAARCGNMGNTNATHATMGNAAYTTNMSNTTYAARTGYWSNGARAGAYTTAAR 393
 406 AAAATCCCGAGTATTTGAAGCAGCGGTATACAGATCTTCAAGAGAGGAGCAAT 465

```

Db 394 ACGTATTCGTCACATTTGGGNGGNGTNGTNGAGTNAATTCNNNGNARCCNGAYGGN 453
OY 466 GATATATCTCTTCAGTGTGGGAGCCCAACCTTTTAAGGACATTCCTGTTAAGAGAA 525
Db 454 AATATGNGNGNTT---YGGNTTGTNCATTTAARATYTTNTNGARCGNGNNAAGCN 510
OY 526 GCTACTGTCTCTGACCTAGAGGCAAGATATTCAACAGGTTTCAGGCCCAAGTGA 585
Db 511 YTNAAAGGNAATGAATGAGATGAATGAAGGNGNAGTNGTNGATGAGGNGTNG 570
OY 586 GCTGAGAGTACTGATCTTCAACAAAAGCCAGTTTATATAGATCCACAGAGAGAA 645
Db 571 GCAAAAGATATATATAGATATACNARMSNGTNGNATGNGARCARARMSNCAY 630
OY 646 GAAATGCTGGAATACATTTGGAATAGGATGAAGTGAAGTGAAGGAGGAGATGCT 705
Db 631 GAWMSNAARCAVACARARMSNGTNGARARARARAGMNGARCARARATATGARGAR 690
OY 706 GATGTACGCTGTGTAGAGGAGCAAGATATTCATGAGTATGACCAATTTTAAGAG 765
Db 691 GARGAARAAAGAYGAYGAYGAYGAYGAYGAYGAYGAYGAYGAYGAYGAYGAYGAR 750
OY 766 CCGAAGAGAGAAAGAAAGAGAGATGCTGCGACGCCAAATGCTCCACAGGAGGTT 825
Db 751 GAYGARGARGARGARATATGARGMSNARGTNACNAACGCTNCARATTCARARAR 810
OY 826 GAGTTTCATTTACCTCTGCTGACCTCCAAAGAGAAAGAAAGAGAGAGAGATG 885
Db 811 GCGNTAAAGMNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGAY 870
OY 886 GGTGAAGAGACCACTATATATGAAATTCCTAAATGCAAGGACATGCAAGAGGT 945
Db 871 GARGARMSNGAYMSNATGAYGAYGAYGAYGAYGAYGAYGAYGAYGAYGAYGAY 930
OY 946 GTAGATCAATTTAATAGAACCAACCACTTAATGAAGAAACAAAGTTCTCTAG 1005
Db 931 GARGARGARGAARAAAGCNGTNCARGTNSNATATATATATATATATATATAT 990
OY 1006 GGCAA 1010
Db 991 GTNAA 995

```

RESULT 5

```

US-08-553-619B-8
; Sequence 8, Application US/08553619B
; Patent No. 5919705
; GENERAL INFORMATION:
; APPLICANT: Dehaan, Petrus T.
; TITLE OF INVENTION: Virus Resistant Plants
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5919705arlis Crop Protection
; STREET: 975 California Avenue
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/553,619B
; FILING DATE: December 1, 1995
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Marcus-Wyner, Lynn
; REGISTRATION NUMBER: 34,869
; REFERENCE/DOCKET NUMBER: 137-1082/PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/354-3588

```

```

; TELEFAX: 415/857-1125
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2621 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; ORIGINAL SOURCE:
; ORGANISM: Chimeric tomato spotted wilt virus S RNA
US-08-553-619B-8

```

```

Query Match 2.5%; Score 42.2; DB 2; Length 2621;
Best Local Similarity 54.1%; Pred. No. 0.094;
Matches 86; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

```

```

OY 1497 TAAGCAAAAAAATCATTTACAGATCTATGAATAGGTAACATTTGAGTAGGTGATTT 1556
Db 667 TATTAATATATTAACCTCTTATCTTATCTTATCTTATCTTATCTTATCTTATCTT 726
OY 1557 AAAATAGTGTGTAATGCAAAATGCTCTATGTTGCTCTGCTCTGCTCTGCTCTGCTCT 1616
Db 727 TAAATATAGTGTGTTTAAATAGTGTGCTTCTGCTGCTGCTGCTGCTTAAATATATAA 786
OY 1617 TAACATATATCTGATGATTAATAAAAAAAAAAAAAAAAAAAAAA 1655
Db 787 TAACAAAAACAAGAAACAAAAAATAATATAATAA 825

```

RESULT 6

```

US-09-461-697-193
; Sequence 193, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COSENT NEUROSCIENCE, Inc.
; APPLICANT: LO, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Putnam, Kasuri
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 193
; LENGTH: 696
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-461-697-193

```

```

Query Match 2.4%; Score 39.6; DB 4; Length 696;
Best Local Similarity 43.6%; Pred. No. 0.25;
Matches 177; Conservative 0; Mismatches 229; Indels 0; Gaps 0;

```

```

OY 517 AAAGGAGAGTACTGCTGCTGACCTAGAGGCAAGATTTCAAAACAGGTTTGACGGC 576
Db 268 AAAGGAGAGTACTGCTGCTGACCTAGAGGCAAGATTTCAAAACAGGTTTGACGGC 327
OY 577 CCAAGTGAAGCTGAGATCTGCTGACCAAAAAAGCCAGTTTATATGAGATCCCA 636
Db 328 AAAATGGAATGGAAGAGAGAGATGCAAAAGAGAAAGATGGAAGAAAGAGTGA 387
OY 637 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 696
Db 388 GAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 447
OY 697 GATGCTGTGATGTCAGCTGTAGAGGCGAGAGATATGATGCTGTGATGATGATTT 756
Db 448 GACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 507

```


TITLE OF INVENTION: CELL DEATH
FILE REFERENCE: 10001-005-999
CURRENT APPLICATION NUMBER: US/09/461,697
CURRENT FILING DATE: 1999-12-14
NUMBER OF SEQ ID NOS: 466
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 187
LENGTH: 774
TYPE: DNA
ORGANISM: Homo sapiens
US-09-461-697-187

Query Match 2.4%; Score 39.6; DB 4; Length 774;
Best Local Similarity 43.6%; Pred. No. 0.26; Mismatches 229; Indels 0; Gaps 0;
Matches 177; Conservative 0;

QY 517 AAAGGAGAGCTACTGCTGCTGACCTAGAAAGCAAGATATCAACAGGGTTGCAGGC 576
DB 346 AAAGGAGAGCTGGAAGAAAGCAAGATATCAACAGGGTTGCAGGC 405
QY 577 CCAAGTGAAGCTGAGAGTACTGCTGACCAAAAGAGGTTATATGAGATCCCA 636
DB 406 AAAATGGAATGAGAAAGAGAGATGCAAAAGAGAGATGAAAAAGTGA 465
QY 637 GAGAGAGAGAAATGCTGAAATACCATTTGGAACCTAGGAGTAACTGCCAAAGAGCA 696
DB 466 GACGAGAAAGAGAAATGAGAAAGAGAGAGAGAGAGATGAAAAAGAGAGAA 525
QY 697 GATGCTGTTGATGTCAGCTTGTAGAGGCGCAACGATATCATGGGTAGTACCAATTTT 756
DB 526 GACAGAGAAAGAGAGAGAGAGAGAGAGATGAGAAAGAGAGAGAGATGAA 585
QY 757 AAGGAGCTCCCTGGAAGAGAGAGAGAGAGAGATGCTGAGCCAAATGCTCACC 816
DB 586 AAGAGGAGAGAGATGTAAGATCAAGAGATGAAAAAGAGAGAGATGAAAAAG 645
QY 817 GGGAGGTTGAGTTTCATTACCTCTCTGACCCCTCAAAAGAGAGAGAGAGAGT 876
DB 646 GATGAAGGTGGAATGAGAGAGAGAGAGAGAGAGATTTTAAAGAGAGAGAA 705
QY 877 AGTGATGACAGCTGAAAGTACCACTATATGAAATTCCTAAAAATG 922
DB 706 GAAGAGAAAGAGAGAGATGATCAAGAGAGATGATGAAAAAG 751

RESULT 10

US-09-461-697-185
Sequence 185, Application US/09461697

Patent No. 6277974

GENERAL INFORMATION:

APPLICANT: COSENT NEUROSCIENCE, Inc.

APPLICANT: Lo, Donald C.

APPLICANT: Barney, Shawn

APPLICANT: Thomas, Mary Beth

APPLICANT: Portbury, Stuart D.

APPLICANT: Puranam, Kasturi

APPLICANT: Katz, Lawrence C.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING

TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING

FILE REFERENCE: 10001-005-999

CURRENT APPLICATION NUMBER: US/09/461,697

CURRENT FILING DATE: 1999-12-14

NUMBER OF SEQ ID NOS: 466

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 185

LENGTH: 819

TYPE: DNA

ORGANISM: Homo sapiens

US-09-461-697-185
Query Match 2.4%; Score 39.6; DB 4; Length 819;
Best Local Similarity 43.6%; Pred. No. 0.27;

Matches 177; Conservative 0; Mismatches 229; Indels 0; Gaps 0;

QY 517 AAAGGAGAGCTACTGCTGCTGACCTAGAAAGCAAGATATCAACAGGGTTGCAGGC 576
DB 391 AAAGGAGAGCTGGAAGAAAGCAAGATGAAAAAGGAGAGAGATGAAAAAGAGAT 450
QY 577 CCAAGTGAAGCTGAGAGTACTGCTGACCAAAAGAGGTTATATGAGATCCCA 636
DB 451 AAAATGGAATGAGAAAGAGAGAGATGCAAAAGAGAGATGAAAAAGTGA 510
QY 637 GAGAGAGAGAAATGCTGAAATACCATTTGGAACCTAGGAGTGAATCTGCCAAAGAGCA 696
DB 511 GACGAGAAAGAGAAATGAGAAAGATGAAAGAGAGAGAGATGAAAAAGAGAGAA 570
QY 697 GATGCTGTTGATGTCAGCTTGTAGAGGCGCAACGATATCATGGGTAGTACCAATTTT 756
DB 571 GACAGAGAAAGAGAGAGAGATGGAAGAGAGATGAGAAAGAGAGAGATGAA 630
QY 757 AAGGAGCTCCCTGGAAGAGAGAGAGAGAGATGCTGAGCCAAATGCTCACC 816
DB 631 AAGAGGAGAGAGATGTAAGATCAAGAGATGAAAAAGAGAGAGATGAAAAAG 690
QY 817 GGGAGGTTGAGTTTCATTACCTCTCTGACCCCTCAAAAGAGAGAGAGAGT 876
DB 691 GATGAAGGTGGAATGAGAGAGAGAGAGAGAGAGATTTTAAAGAGAGAGAA 750
QY 877 AGTGATGACAGCTGAAAGTACCACTATATGAAATTCCTAAAAATG 922
DB 751 GAAGAGAAAGAGAGAGATGATCAAGAGAGATGATGAAAAAG 796

RESULT 11

US-09-461-697-184
Sequence 184, Application US/09461697

Patent No. 6277974

GENERAL INFORMATION:

APPLICANT: COSENT NEUROSCIENCE, Inc.

APPLICANT: Lo, Donald C.

APPLICANT: Barney, Shawn

APPLICANT: Thomas, Mary Beth

APPLICANT: Portbury, Stuart D.

APPLICANT: Puranam, Kasturi

APPLICANT: Katz, Lawrence C.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING

TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING

FILE REFERENCE: 10001-005-999

CURRENT APPLICATION NUMBER: US/09/461,697

CURRENT FILING DATE: 1999-12-14

NUMBER OF SEQ ID NOS: 466

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 184

LENGTH: 1669

TYPE: DNA

ORGANISM: Homo sapiens

US-09-461-697-184

Query Match 2.4%; Score 39.6; DB 4; Length 1669;
Best Local Similarity 43.6%; Pred. No. 0.38;
Matches 177; Conservative 0; Mismatches 229; Indels 0; Gaps 0;

QY 517 AAAGGAGAGCTACTGCTGCTGACCTAGAAAGCAAGATATCAACAGGGTTGCAGGC 576
DB 470 AAAGGAGAGCTGGAAGAAAGCAAGATGAAAAAGGAGAGAGATGAAAAAGAGAT 529
QY 577 CCAAGTGAAGCTGAGAGTACTGCTGACCAAAAGAGGTTATATGAGATCCCA 636
DB 530 AAAATGGAATGAGAAAGAGAGAGATGCAAAAGAGAGAGATGAAAAAGTGA 589
QY 637 GAGAGAGAGAAATGCTGAAATACCATTTGGAACCTAGGAGTGAATCTGCCAAAGAGCA 696
DB 590 GACGAGAAAGAGAAATGAGAAAGATGGAAGAGAGAGAGATGAAAAAGAGAGAA 649

RESULT 12
US-08-714-918-11/c
; Sequence 11, Application US/08714918

Query Match	2.4%	Score 39.2;	DB 3;	Length 719;
Best Local Similarity	56.1%	Pred. NO. 0.33;		
Matches 74;	Conservative	0;	Mismatches 58;	Indels 0;
				Gaps 0;

RESULT 13
US-09-265-315-11/c
; Sequence 11, Application US/09265315
; Patent No. 6187541

Query Match	2.48;	Score 39.2;	DB 4;	Length 719;
Best Local Similarity	56.18;	Pred. No. 0.33;		
Matches 74; Conservative	0;	Mismatches 58;	Indels 0;	Gaps 0;

QY 1489 AATGTTTTCAGCAAAAAAATCATTCATTCAGATCTATGAAATNGATACATTGGAGTAG 1548
 Db 378 AACTATGTTTCATTCACAACAATTCATTCATTTCGCAACTGTGAATATGTCATATATTTGCTTTG 319
 QY 1549 TGTCAATTAAAAATAGTGTGATGATGTCACAAATGCGTCTCTATGTTGTTGCTGTGAGA 1608
 Db 318 GAGATATAAAAAGCACTGTGCAAAAACAAGTGTCTTAACTAATTAATTTGTTCACT 259
 QY 1609 CATGAAATATAA 1620
 Db 258 AGTTTGAATCAA 247

```

1      RESULT 15
2      US-09-266-417-11/c
3      ; Sequence 11, Application US/09266417
4      ; Patent No. 6228588
5      GENERAL INFORMATION:
6      ; APPLICANT: Benton, Bret
7      ; APPLICANT: Lee, Ving J
8      ; APPLICANT: Malouin, Francois
9      ; APPLICANT: Martin, Patrick K.
10     ; APPLICANT: Schmid, Molly B.
11     ; APPLICANT: Sun, Dongxu
12     ; TITLE OF INVENTION: METHODS OF SCREENING FOR COMPOUNDS
13     ; TITLE OF INVENTION: ACTIVE ON STAPHYLOCOCCUS AUREUS
14     ; TITLE OF INVENTION: TARGET GENES
15     ; NUMBER OF SEQUENCES: 111
16     CORRESPONDENCE ADDRESS:
17     ; ADDRESSEE: Lyon & Lyon
18     ; STREET: 633 West Fifth Street
19     ; STREET: Suite 4700
20     ; CITY: Los Angeles
21     ; STATE: California
22     ; COUNTRY: U.S.A.
23     ; ZIP: 90071-2066
24     COMPUTER READABLE FORM:
25     ; MEDIUM TYPE: 3.5" diskette, 1.44 Mb
26     ; MEDIUM TYPE: storage
27     ; COMPUTER: IBM Compatible
28     ; OPERATING SYSTEM: IBM P.C. DOS 5.0
29     ; SOFTWARE: Word Perfect 5.1
30     CURRENT APPLICATION DATA:
31     ; APPLICATION NUMBER: US/09/266,417
32     ; FILING DATE: March 9, 1999
33     ; CLASSIFICATION: 435
34     PRIOR APPLICATION DATA:
35     ; APPLICATION NUMBER: 08/714,918
36     ; FILING DATE: September 13, 1996
37     ; APPLICATION NUMBER: 60/009,102
38     ; FILING DATE: December 22, 1995
39     ; APPLICATION NUMBER: 60/003,798
40     ; FILING DATE: September 15, 1995
41     ATTORNEY/AGENT INFORMATION:
42     ; NAME: Waiburg, Richard J.
43     ; REGISTRATION NUMBER: 32,327
44     ; REFERENCE/DOCKET NUMBER: 240/248
45     TELECOMMUNICATION INFORMATION:
46     ; TELEPHONE: (213) 489-1600
47     ; TELEFAX: (213) 955-0440
48     ; TELEX: 67-3510
49     INFORMATION FOR SEQ. ID NO.: 11:
50     ; SEQUENCE CHARACTERISTICS:
51     ; LENGTH: 719 base pairs
52     ; TYPE: nucleic acid
53     ; STRANDEDNESS: single
54     ; TOPOLOGY: linear
55     US-09-266-417-11

```

Query Match	2.48;	Score 39.2;	DB 4;	Length 719;
Best Local Similarity	56.18;	Pred. No. 0.33;		
Matches 74; Conservative	0;	Mismatches 58;	Indels 0;	Gaps 0

us-09-700-696b-1.rni

Page 9

Search completed: April 16, 2003, 19:47:52
Job time : 114 secs

GenCore version 5.1.4.P5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 16, 2003, 15:20:04 ; Search time 198 Seconds
(without alignments)
8396.717 Million cell updates/sec

Title: US-09-700-696B-1

Perfect score: 1655

Sequence: 1 gtgataaagaatagatagat.....ataaaaaaaaaaaaaaaaaa 1655

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 639749 seqs, 502280978 residues

Total number of hits satisfying chosen parameters: 1279498

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications_NA:*

- 1: /cgn2_6/ptodata/2/pubna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubna/PCIT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubna/US07_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubna/PCITUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/2/pubna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	45.8	2.8	2000	US-09-887-576-831	Sequence 831, App
C 2	43.6	2.6	341	US-09-960-352-12302	Sequence 12302, A
C 3	42.8	2.6	32195	US-10-102-627-92	Sequence 92, Appl
C 4	42.6	2.6	747	US-10-123-155-212	Sequence 212, App
C 5	42.6	2.6	3986	US-10-087-464-42	Sequence 42, Appl
C 6	42.2	2.5	529	US-09-983-965-2109	Sequence 2109, App
C 7	42.2	2.5	1959	US-09-864-761-4012	Sequence 4012, A
C 8	41.2	2.5	424	US-09-960-352-11218	Sequence 11218, A
C 9	41	2.5	545	US-09-878-574-4299	Sequence 4299, App
C 10	41	2.5	1000	US-09-938-8428-3924	Sequence 3924, App
C 11	40.8	2.5	516	US-09-960-352-5785	Sequence 5785, App
C 12	40.2	2.4	2000	US-09-938-8428-3607	Sequence 3607, App
C 13	40	2.4	396	US-09-970-966-53	Sequence 53, Appl
C 14	40	2.4	396	US-09-825-294-53	Sequence 53, Appl
C 15	39.8	2.4	321	US-09-960-352-1267	Sequence 1267, App
C 16	39.6	2.4	380	US-09-960-352-3640	Sequence 3640, App
C 17	39.6	2.4	686	US-09-922-261-193	Sequence 193, App
C 18	39.6	2.4	699	US-09-922-261-191	Sequence 191, App
C 19	39.6	2.4	717	US-09-922-261-189	Sequence 189, App

20	39.6	2.4	774	10	US-09-922-261-187	Sequence 187, App
21	39.6	2.4	819	10	US-09-922-261-185	Sequence 185, App
22	39.6	2.4	1669	10	US-09-922-261-184	Sequence 184, App
C 23	39.4	2.4	381	10	US-09-960-352-13840	Sequence 13840, A
C 24	39.2	2.4	1811	10	US-10-086-510-2	Sequence 2, Appl
C 25	39.2	2.4	6645	7	US-08-781-986A-259	Sequence 259, Appl
C 26	39	2.4	32035	9	US-10-092-154-1611	Sequence 1611, App
C 27	39	2.4	32035	10	US-09-764-847-1611	Sequence 1611, App
C 28	39	2.4	640681	10	US-09-790-988-1	Sequence 1, Appl
C 29	38.8	2.3	379	10	US-09-960-352-3105	Sequence 3105, App
C 30	38.8	2.3	1940	9	US-09-798-889-18	Sequence 18, Appl
C 31	38.6	2.3	570	10	US-09-864-761-21778	Sequence 21778, A
C 32	38.6	2.3	1360	9	US-09-895-913A-151	Sequence 151, App
C 33	38.6	2.3	1976	10	US-09-864-761-5052	Sequence 5052, App
C 34	38.4	2.3	419	10	US-09-960-352-11234	Sequence 11234, A
C 35	38.2	2.3	314	9	US-09-954-531-1122	Sequence 1122, App
C 36	38.2	2.3	520	9	US-10-184-644-332	Sequence 332, App
C 37	38.2	2.3	520	9	US-10-184-634-332	Sequence 332, App
C 38	38	2.3	446	10	US-09-960-352-3400	Sequence 3400, App
C 39	38	2.3	609	10	US-09-770-149-886	Sequence 886, App
C 40	38	2.3	1267	12	US-10-001-843-45	Sequence 45, Appl
C 41	38	2.3	1944	10	US-09-864-761-2825	Sequence 2825, App
C 42	38	2.3	2007	10	US-09-452-599-169	Sequence 169, App
C 43	38	2.3	16428	9	US-10-092-154-1613	Sequence 1613, App
C 44	38	2.3	16428	10	US-09-764-847-1613	Sequence 1613, App
C 45	38	2.3	31834	9	US-10-092-154-1612	Sequence 1612, App

ALIGNMENTS

RESULT 1
US-09-887-576-831/C
Sequence 831, Application US/09887576
Patent No. US20020144047A1
GENERAL INFORMATION:
APPLICANT: Budworth, P.
APPLICANT: Brown, D.
APPLICANT: Chang, H.
APPLICANT: Zhu, T.
APPLICANT: Han, B.
APPLICANT: Wang, X.
APPLICANT: Cooper, Bret
TITLE OF INVENTION: Promoters for regulation of plant expression
FILE REFERENCE: 1360.001US1
CURRENT APPLICATION NUMBER: US/09/887,576
CURRENT FILING DATE: 2001-06-25
PRIOR APPLICATION NUMBER: US 60/213,848
PRIOR FILING DATE: 2000-06-23
PRIOR APPLICATION NUMBER: US 60/214,087
PRIOR FILING DATE: 2000-06-23
PRIOR APPLICATION NUMBER: US 60/258,692
PRIOR FILING DATE: 2000-12-29
NUMBER OF SEQ ID NOS: 875
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 831
LENGTH: 2000
TYPE: DNA
ORGANISM: Oryza sativa
US-09-887-576-831

Query Match 2.8%; Score 45.8; DB 10; Length 2000;
Best Local Similarity 53.0%; Pred. No. 0.11;
Matches 98; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 1471 AATTCATTAAGGCTATATGTTTAAAGCAAAAAAATCATTCAGATCTATGAAT 1530
DB 1269 ACTACACGCTGATATCTGACTGTTACCAATATTTACTCCATCTTAAAGA 1210
QY 1531 AGGTACATTTGGAGTGTGCTCTTAAATAATAGTTGGAATGCACAATGCCCTTCA 1590
DB 1209 AAGTATTTAGGCAATGCTTAATCAATCTTAGGAATATATCAATTAATTAATCTTTTA 1150

QY	194	CGATTTAAC-TCCTGGGGGAGAAAAACAAGGAACACACCTGAGATGCTCTAAACCTA	252
Db	1712	ATATTAAAAAGTGTACTAGTAATTAATTAACCTATGATPAACAATATGATATGTAACGAA	1771
QY	253	ATCCCAACAGTAT-----GAATTATGCTAAAGCAGCTCGAAGGATTAAGAAGC	304
Db	1772	GTAATGACATTTAATTAATGAGGTATATGATGATGACAGTAAATTAATGATTTAATA	1831
QY	305	CTCAAGAGATTCACAGCCCAAGAAAGTCAGATAAAGCAAAAGCACCCTCTATTC	364
Db	1832	GTAATAGAGCTATTAATAGTAATTAATATGAGATGATTAATTAACATATATTAATATG	1891
QY	365	AACACACATTGACTACCTAAACATCTCTCAAAAGTCAAAAAATCCCACTGATTTG	424
Db	1892	ATGGAATGTGATTAATTAATTAATTAACCTAATTAATTAATTAATTAATTAATTT	1951
QY	425	AAGGAGGCGGTATACAGATCTTCAAGAGAGAGGGGACATGATATATCTCCTTCAGTG	484
Db	1952	ATGTAACAAATTAATTAATTAACAACAATTAATTAAGGACATTAATTAATGATGGAATG	2011
QY	485	GGGACGCCAACCTTTAAGACACTTCCTGGTAAAGAGAGCTACTGGCTGACCTTG	544
Db	2012	GATAGTAATTAATTAATTAATTAATGATGATGATGACGAAGAAGAAGAGATGAGGAAGT	2071
QY	545	AAGGCAAAGATATTCACAAAGGGTTGCAAGGCCCACTGAAGCTGAGAGTACTCATCTTG	604
Db	2072	ATTAACATTAATTAATTAATGATGATGATTAATTAAGCTGATTAACGAAGAATGGAAGTAATG	2131
QY	605	ACACAAAAAGCAGCTTATATGATCCAGAGAGAGAAAGAAATGTGTGAAA	659
Db	2132	ATGAAGATACCATAGTATTAATTAATGATTAATGATTAATTAATGAGAGAAA	2186

```

RESULT 6
US-09-983-965-2109/c
Sequence 2109, Application US/09983965
Patent No. US20020137160A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Nengbing
APPLICANT: Byatt, John C.
APPLICANT: Mahalingam, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND FAT DEPOSITION
TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
FILE REFERENCE: 37-21(10297)C
CURRENT APPLICATION NUMBER: US/09/983, 965
CURRENT FILING DATE: 2001-10-26
PRIOR APPLICATION NUMBER: US 09/465,231
PRIOR FILING DATE: 1999-12-15
PRIOR APPLICATION NUMBER: US 60/113,678
PRIOR FILING DATE: 1998-12-17
NUMBER OF SEQ ID NOS: 5912
SEQ ID NO 2109
LENGTH: 529
TYPE: DNA
ORGANISM: Bos taurus
FEATURE:

```

Query Match	2.5%	Score 42.2	DB 10	Length 529
Best Local Similarity	43.9%	Pred. No. 0.4		
Matches 179	Conservative	0	Mismatches 229	Indels 0
Gaps				0
QY	4	AAATTAACATATATGTCTTCAGTAAACAAGATCTACACATGGCCCTGAGAGATGTCAATT	63	
Db	489	AA	430	
QY	64	TATCTTAAGTCACACTGGGAATTAAGGCTTTGAGATGAGATGATGCTATTCAGCAACTA	123	

```

Db 429 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 370
QY 124 CATGACCAAGAGATATGCGCAGCTCTCATCAGAAATTAACATCATATATATGGG 183
Db 369 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 310
QY 184 CCAGTACTGCATTAATCTCTCTGGGAGAGAAAGAACACACTAGATGTT 243
Db 309 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 250
QY 244 CTAAACATTAATCCAGCAATGATGATATGCTTAAGACACTGAGATATAAAGAG 303
Db 249 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 190
QY 304 CCTCAAGAGATTCCTCAAGCCCAAGAAAGTCAATAAAGCAAGACCCATGATTT 363
Db 189 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 130
QY 364 CAACACACATTTGACTACTCAATAAATCTCTCAAAAGTCAAAATATC 411
Db 129 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 82

```

RESULT 7

```

US-09-864-761-4012
; Sequence 4012, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecm1ca-x-1
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203

```

```

; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 4012
; LENGTH: 1959
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL022334.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.3
US-09-864-761-4012

```

Query Match 2.5%; Score 42; DB 10; Length 1959;
 Best Local Similarity 42.7%; Pred. No. 1;
 Matches 274; Conservative 0; Mismatches 365; Indels 3; Gaps 1;

```

QY 324 CCAGAAAGTCCAGTAAAGCAAGCAAGCCATGATTCACCAACATGACTACT 383
Db 326 CAATTAACACCTAAAGAAAGCAAGAGGCTATGGCTTGAGACATGATGATGCA 385
QY 384 AAACATCTCTCAAAAGCAAAAGCAAAAGCCAGATTTTGAAGCCAGCGTTATACGA 443
Db 386 CAATGATGCTGCTATGATGATGATGATGATGATGATGATGATGATGATGATG 445
QY 444 TCTTCAAGAGAGAGGAGCAATGATATATCTCTTCAAGAGAGAGAGAGAGAGAG 503
Db 446 TAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 505
QY 504 GGACATTCCTGCTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 563
Db 506 GCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 565
QY 564 AGGTTTTCAGAGCCCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 623
Db 566 TGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 625
QY 624 TAATGATATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 683
Db 626 TGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 682
QY 684 TGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 743
Db 683 TGATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 742
QY 744 TACTACCAATTTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 803
Db 743 TGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 802
QY 804 AATGCTCACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 863
Db 803 TGATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 862
QY 864 AAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 923
Db 863 TGATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 922
QY 924 CAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 965
Db 923 TGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 964

```

RESULT 8
 US-09-960-352-11218
 ; Sequence 11218, Application US/09960352

Query Match	2.5%	Score 41;	DB 10;	Length 545;
Best Local Similarity	50.0%	Pred. No. 0.88;	Indels 0;	Gaps
Matches 101;	Conservative			

0y	208	GGGGAAGAAAACAAGACGACCTAGGAATGTTTAAACATTAATCCGCAAGATATG	267

```

RESULT 11
US-09-960-352-5785/c
; Sequence 5785, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Ningbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mahalingam, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND

```

;; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
;; FILE REFERENCE: 16511.006/37-21(10298)C
;; CURRENT APPLICATION NUMBER: US/09/960,352
;; CURRENT FILING DATE: 2001-09-24
;; NUMBER OF SEQ ID NOS: 15112
;; SEQ ID NO 5785
;; LENGTH: 516
;; TYPE: DNA
;; ORGANISM: Bos taurus
;; FEATURE:
;; NAME/KEY: unsure
;; LOCATION: (76) . (90)
;; OTHER INFORMATION: unsure at all n locations
;; OTHER INFORMATION: Clone ID: 25-LIB3057-018-Q1-K1-G1
US-09-960-352-5785

Query Match 2.5%; Score 40.8; DB 10; Length 516;
Best Local Similarity 46.2%; Pred. No. 0.96;
Matches 135; Conservative 0; Mismatches 157; Indels 0; Gaps 0;

QY 118 AACTTACATGACCAAGAAATATGGCGAGCTCTCATGCAAAATACATGCAATATA 177
DB 432 AAAAAATATAAAAAATATAAAAAATATAAAAAATATAAAAAATATAAAAAAT 373
QY 178 ATGGGGCGAGTGCCTTAATCTCTGGGGAAGAAACAAAGACACACCTAGG 237
DB 372 ATATAAAAAATATAAAAAATATAAAAAATATAAAAAATATAAAAAATATA 313
QY 238 AATGTCATACATATATCCAGCAAGTATGATATGCTAAAGCAGCTCGAAGATATA 297
DB 312 AAAAAATATAAAAAATATAAAAAATATAAAAAATATAAAAAATATAAAAAAT 253
QY 298 AAGAAGCCCAAGAGATTCCCAAGCCCAAGAAAGTCAGTAAAGCAAGCACCAT 357
DB 252 AAAAAATATAAAAAATATAAAAAATATAAAAAATATAAAAAATATAAAAAAT 193
QY 358 CGTATTCACACATGACTGACTACTTAACATCTCTCAAAAGTCAGCAAAAA 409
DB 192 AAAAAATATAAAAAATATAAAAAATATAAAAAATATAAAAAATATAAAAAAT 141

RESULT 12
US-09-938-842A-3607
; Sequence 3607, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 3607
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-3607

Query Match 2.4%; Score 40.2; DB 9; Length 2000;
Best Local Similarity 53.2%; Pred. No. 3.1;
Matches 107; Conservative 0; Mismatches 93; Indels 1; Gaps 1;
QY 1454 TAATAGTCACAGTATAAAATCTATTAAAGGCTATATGTTTAAAGCAAAAAAAATCA 1513

DB 1521 TAAGATCTCATATAATACATTTTAAAGCAGATATTAATTTTACAAATGAAGAAAA 1580
QY 1514 TTACAGATCTATGAATAGTAACTTTGAGAGTGCATTTAAATAGTGGTGAAT 1573
DB 1581 ATATGAATCTATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1640
QY 1574 GTACAAATGCTCTATGTTGTTGCTGTAGACGAAATAAACAATATCTGCA 1633
DB 1641 GTCAATAT - CTTTATTTATTTGCTGTGCAAACTATGTGAATTAATCTCGAG 1699
QY 1634 TGATAAAAAATAAAAA 1654
DB 1700 AATATATTGAACAAAAA 1720

RESULT 13
US-09-970-966-53/C
; Sequence 53, Application US/09970966
; Patent No. US20020173638A1
; GENERAL INFORMATION:
; APPLICANT: Stolk, John A.
; APPLICANT: Molesh, David Alan
; APPLICANT: Fling, Steven P.
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.484C6
; CURRENT APPLICATION NUMBER: US/09/970,966
; CURRENT FILING DATE: 2001-10-02
; NUMBER OF SEQ ID NOS: 215
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 53
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 224, 225, 228, 235, 240, 246, 257, 266, 274, 279, 281, 282,
; LOCATION: 283, 285, 287, 288, 290, 291, 292, 293, 294, 295, 296, 297,
; LOCATION: 300, 301, 303, 307, 311, 313, 314, 317, 318, 319, 320, 321,
; LOCATION: 323, 324, 328, 329, 330, 336, 337, 338, 339, 340, 341
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc.feature
; LOCATION: 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 356,
; LOCATION: 357, 358, 359, 362, 363, 364, 365, 366, 367, 373, 380, 381,
; LOCATION: 382, 385, 387, 388, 389, 390, 392
; OTHER INFORMATION: n = A,T,C or G
US-09-970-966-53

Query Match 2.4%; Score 40; DB 9; Length 396;
Best Local Similarity 36.9%; Pred. No. 1.3;
Matches 121; Conservative 0; Mismatches 207; Indels 0; Gaps 0;

QY 82 AATAAGGTTTGAGATGAGATGATGCTATCAGCAACTCATGACCAAGAAATAT 141
DB 383 ANNNAAGGCGCAGGNNNNNNNNAAGANNNNNNNNNNNNNNNNNNNNNNNNNNCCN 324
QY 142 GGGCAGCTCTCATGAGAAATACATGCAACATATAATGGGCGCTAGTGCATTAA 201
DB 323 NTRNNNNNANNANTATNAAGNANNATNNNNNNNNNNNNNTNTNTTNTTTCTTWT 264
QY 202 CTCCTGGGGGAAGAAAGAGACACACCTAGCATGTTCTTAACATATCCGCA 261
DB 263 CTTTTTNGAATTAAGAAAGNAAAAAANNNTTAAAAAATAAAAA 204
QY 262 AGTATGATTTATGTTAAAGCACTCGAAGATATAAAGAAAGCTCAAGAGATTCCCA 321
DB 203 AAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAA 144
QY 322 GCCCGAAAGCTCGTAAAAAGCAAGCAGCATGCTATTCACACCAATGACTAC 381
DB 143 AAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAA 84

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2003, 11:56:31 ; Search time 76 Seconds

(without alignments)
753.918 Million cell updates/sec

Title: US-09-700-696b-2

Perfect score: 2279

Sequence: 1 VNKEYSISNKNENTHNGLRMS.....RRDSSSESDSGSSSESDDG 430

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_101002:*

1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*

2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*

3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*

4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*

5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*

6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*

7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*

8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*

9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*

10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*

11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*

12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*

13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*

14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*

15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*

16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*

17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*

18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*

19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*

20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*

21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*

22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the total score distribution, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2279	100.0	430	21	AAV53812	Amino acid sequenc
2	2279	100.0	430	22	AAV53812	truncated phosphat
3	2276	99.9	509	22	AAV53812	Human osteoregulin
4	2276	99.9	525	22	AAV53812	Human osteoregulin
5	2276	99.9	525	22	AAV53812	Human osteoregulin
6	2276	99.9	525	22	AAV53812	Human osteoregulin
7	2276	99.9	540	22	AAV53812	Human osteoregulin
8	2276	99.9	556	22	AAV53812	Human osteoregulin
9	2269	99.6	525	23	AAV53812	Human osteoregulin
10	983.5	43.2	441	22	AAV53812	Mouse osteoregulin

11	938.5	41.2	435	22	AAV53812	Rat osteoregulin.
12	840	36.9	275	22	AAV53812	Human polypeptide.
13	514	22.6	97	23	AAV53812	Dental product bon
14	514	22.6	97	23	AAV53812	Dental product bon
15	246	10.8	47	23	AAV53812	Dental product bon
16	246	10.8	47	23	AAV53812	Dental product bon
17	235.5	10.3	47	23	AAV53812	Dental product bon
18	235.5	10.3	47	23	AAV53812	Dental product bon
19	229	10.0	47	23	AAV53812	Dental product bon
20	229	10.0	47	23	AAV53812	Dental product bon
21	218	9.6	40	23	AAV53812	Dental product bon
22	218	9.6	40	23	AAV53812	Dental product bon
23	216	9.5	44	23	AAV53812	Dental product bon
24	216	9.5	44	23	AAV53812	Dental product bon
25	215	9.4	44	23	AAV53812	Dental product bon
26	215	9.4	44	23	AAV53812	Dental product bon
27	214.5	9.4	45	23	AAV53812	Dental product bon
28	214.5	9.4	45	23	AAV53812	Dental product bon
29	198.5	8.7	41	23	AAV53812	Dental product bon
30	198.5	8.7	41	23	AAV53812	Dental product bon
31	190.5	8.4	40	23	AAV53812	Dental product bon
32	190.5	8.4	40	23	AAV53812	Dental product bon
33	188.5	8.3	40	23	AAV53812	Dental product bon
34	188.5	8.3	40	23	AAV53812	Dental product bon
35	182	8.0	37	23	AAV53812	Dental product bon
36	182	8.0	37	23	AAV53812	Dental product bon
37	180	7.9	38	23	AAV53812	Dental product bon
38	180	7.9	38	23	AAV53812	Dental product bon
39	165	7.2	35	23	AAV53812	Dental product bon
40	163.5	7.2	35	23	AAV53812	Dental product bon
41	163.5	7.2	35	23	AAV53812	Dental product bon
42	157.5	6.9	35	23	AAV53812	Dental product bon
43	157.5	6.9	35	23	AAV53812	Dental product bon
44	153.5	6.7	33	23	AAV53812	Dental product bon
45	153.5	6.7	33	23	AAV53812	Dental product bon

ALIGNMENTS

RESULT 1

AAV53812

AAV53812 standard; Protein: 430 AA.

AC XX

AC XX

DT 22-FEB-2000 (first entry)

Amino acid sequence of a human phosphatonin polypeptide.

Human: phosphatonin; Metastatic-tumour Excreted Phosphaturic-Element;
MEPP: Na+-dependent phosphate cotransport; vitamin D metabolism;
bone mineralization; phosphate metabolism related disease;
hyperphosphatemia; renal osteodystrophy; renal dialysis;
secondary hyperparathyroidism; osteitis fibrosa cystica; hypercalcaemia;
X-linked hypophosphatemic rickets; hereditary hypophosphatemic rickets;
hypomineralized bone lesion; stunted growth; cystic fibrosis;
oncogenic hypophosphatemic osteomalacia; renal phosphate leakage;
renal osteodystrophy; osteoporosis; vitamin D resistant rickets;
end organ resistance; renal Fanconi syndrome; autosomal rickets;
Paget's disease; kidney failure; renal tubular acidosis; sprue.

Homo sapiens.

Key Location/Qualifiers

Modified-site 8..10 "protein kinase C phosphorylation site"

Modified-site 8..11 "Casein kinase II phosphorylation site"

Modified-site 16..21 "myristoylation site"

Modified-site 40..47 "tyrosine kinase phosphorylation site"

FT Modified-site /note- "Tyrosine kinase phosphorylation site"

FT Modified-site 77..79
 FT /note="protein kinase C phosphorylation site"
 FT Modified-site 118..120
 FT /note="protein kinase C phosphorylation site"
 FT Modified-site 119..224
 FT /note="myristoylation site"
 FT Modified-site 139..142
 FT /note="Casein kinase II phosphorylation site"
 FT Modified-site 143..148
 FT /note="myristoylation site"
 FT Region 152..154
 FT /note="cell attachment tripeptide"
 FT Region 161..165
 FT /note="glycosaminoglycan attachment site"
 FT Modified-site 177..180
 FT /note="Casein kinase II phosphorylation site"
 FT Modified-site 194..197
 FT /note="Casein kinase II phosphorylation site"
 FT Modified-site 199..202
 FT /note="Casein kinase II phosphorylation site"
 FT Modified-site 203..205
 FT /note="protein kinase C phosphorylation site"
 FT Modified-site 224..227
 FT /note="Casein kinase II phosphorylation site"
 FT Modified-site 228..231
 FT /note="Casein kinase II phosphorylation site"
 FT Modified-site 228..230
 FT /note="protein kinase C phosphorylation site"
 FT Modified-site 238..241
 FT /note="Casein kinase II phosphorylation site"
 FT Modified-site 266..271
 FT /note="myristoylation site"
 FT Modified-site 291..296
 FT /note="myristoylation site"
 FT Modified-site 311..313
 FT /note="protein kinase C phosphorylation site"
 FT Modified-site 312..314
 FT /note="protein kinase C phosphorylation site"
 FT Modified-site 315..320
 FT /note="myristoylation site"
 FT Modified-site 319..321
 FT /note="protein kinase C phosphorylation site"
 FT Modified-site 325..328
 FT /note="Casein kinase II phosphorylation site"
 FT Modified-site 370..373
 FT /note="amidation site"
 FT Modified-site 382..386
 FT /note="Asu-glycosylation site"
 FT Modified-site 383..387
 FT /note="Asu-glycosylation site"
 FT Modified-site 384..386
 FT /note="protein kinase C phosphorylation site"
 FT Modified-site 389..394
 FT /note="myristoylation site"
 FT Modified-site 403..405
 FT /note="protein kinase C phosphorylation site"
 FT Modified-site 405..408
 FT /note="cAMP and cGMP dependent protein kinase phosphorylation site"
 FT Modified-site 408..410
 FT /note="protein kinase C phosphorylation site"
 FT Modified-site 409..411
 FT /note="protein kinase C phosphorylation site"
 FT Modified-site 423..426
 FT /note="Casein kinase II phosphorylation site"
 FT Modified-site 425..428
 FT /note="Casein kinase II phosphorylation site"
 FT Modified-site 427..430
 FT /note="Casein kinase II phosphorylation site"
 XX
 PN MO9960017-A2.
 XX
 PD 25-NOV-1999.

XX PF 18-MAY-1999; 99WO-EP03403.
 XX PR 18-MAY-1998; 98GB-0010681.
 XX PR 04-SEP-1998; 98GB-0019387.
 XX (UNLO) UNIV COLLEGE LONDON.
 XX
 XX PI Rowe P;
 XX DR WPI; 2000-053262/04.
 XX DR N-PDB; AAZ36447.
 PT New polypeptides involved in the regulation of phosphate metabolism
 PT useful for diagnosing and treating disorders related to phosphate
 PT metabolism
 PS
 XX
 CC Claim 6; Fig 8; 136pp; English.
 CC
 CC The present sequence represents a phosphatonin polypeptide (also called
 CC Metastatic-tumour Excreted Phosphaturic-Element (MEPE)). The level of
 CC phosphatonin in a subject modulates Na⁺-dependent phosphate cotransport,
 CC vitamin D metabolism and/or bone mineralisation. The phosphatonin
 CC polypeptides, polynucleotides, vectors and antibodies are used to treat
 CC phosphate metabolism related disease. They are used for treatment of
 CC hyperphosphatemia, or renal osteodystrophy, hyperphosphatemia in renal
 CC dialysis/pre-dialysis, secondary hyperparathyroidism or osteitis fibrosa
 CC cystica, or X-linked hypophosphatemic rickets, hereditary
 CC hypophosphatemic rickets with hypercalcaemia (HHR), hypomineralised
 CC bone lesions, stunted growth in juveniles, oncogenic hypophosphatemic
 CC osteomalacia, renal phosphate leakage, renal osteodystrophy,
 CC Fanconi syndrome, autosomal rickets, Paget's disease, kidney failure,
 CC renal tubular acidosis, cystic fibrosis or sprue. The polypeptide
 CC may also be used to manufacture combined preparations for simultaneous
 CC separate or sequential use for the treatment of phosphate metabolism
 CC disorders. A transformed osteoblast or bone cell line capable of
 CC phosphatonin overexpression is useful for the production of
 CC phosphatonin.
 CC
 CC
 SO Sequence 430 AA;
 Query Match 100.0%; Score 2279; DB 21; Length 430;
 Best Local Similarity 100.0%; Pred. No. 1.7e-181;
 Matches 430; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VNKEYSISNKENHTNGILMSITYPKSTGNKGFEDGDDAISKLHDEEFGALIRNNMHIM 60
 DB 1 VNKEYSISNKENHTNGILMSITYPKSTGNKGFEDGDDAISKLHDEEFGALIRNNMHIM 60
 QY 61 GPVTAIKLLGEENKENTFRNVLTITPASMNYAKAKSKKKRPQDSQAOKSPVSKSTHR 120
 DB 61 GPVTAIKLLGEENKENTFRNVLTITPASMNYAKAKSKKKRPQDSQAOKSPVSKSTHR 120
 QY 61 GPVTAIKLLGEENKENTFRNVLTITPASMNYAKAKSKKKRPQDSQAOKSPVSKSTHR 120
 DB 61 GPVTAIKLLGEENKENTFRNVLTITPASMNYAKAKSKKKRPQDSQAOKSPVSKSTHR 120
 QY 121 IOHNIDYLLKLSKVKKIPDSFGSGYTDLOERGDNDISPFSGDQGFKDIPGKBATGPD 180
 DB 121 IOHNIDYLLKLSKVKKIPDSFGSGYTDLOERGDNDISPFSGDQGFKDIPGKBATGPD 180
 QY 121 IOHNIDYLLKLSKVKKIPDSFGSGYTDLOERGDNDISPFSGDQGFKDIPGKBATGPD 180
 DB 121 IOHNIDYLLKLSKVKKIPDSFGSGYTDLOERGDNDISPFSGDQGFKDIPGKBATGPD 180
 QY 181 LEGKDIOGTGAPSAESTHLDTKKPGYNEIPERENGNTTGTDETAKEADAVDSLV 240
 DB 181 LEGKDIOGTGAPSAESTHLDTKKPGYNEIPERENGNTTGTDETAKEADAVDSLV 240
 QY 241 EGSNDIMGSTNFKELPGRGNRVVDAGSONAHOGKVEFHPYPAPSKKKRKSDDAESTN 300
 DB 241 EGSNDIMGSTNFKELPGRGNRVVDAGSONAHOGKVEFHPYPAPSKKKRKSDDAESTN 300
 QY 241 EGSNDIMGSTNFKELPGRGNRVVDAGSONAHOGKVEFHPYPAPSKKKRKSDDAESTN 300
 DB 241 EGSNDIMGSTNFKELPGRGNRVVDAGSONAHOGKVEFHPYPAPSKKKRKSDDAESTN 300
 QY 301 YNEIPKNGKSTRKGVDSNRQAATLNEKORPPSKGSGGLPIPRGLDNEIKNMDSEFN 360
 DB 301 YNEIPKNGKSTRKGVDSNRQAATLNEKORPPSKGSGGLPIPRGLDNEIKNMDSEFN 360
 QY 361 GPSHENITTHGRKRYHVPKRONNSTRNKMGPGKSGSWGROPNSNRFSRRDDSSSSD 420
 DB 361 GPSHENITTHGRKRYHVPKRONNSTRNKMGPGKSGSWGROPNSNRFSRRDDSSSSD 420
 QY 361 GPSHENITTHGRKRYHVPKRONNSTRNKMGPGKSGSWGROPNSNRFSRRDDSSSSD 420
 DB 361 GPSHENITTHGRKRYHVPKRONNSTRNKMGPGKSGSWGROPNSNRFSRRDDSSSSD 420

QY 421 SSSSESDGD 430
 |||||
 Db 421 SSSSESDGD 430

RESULT 2
 AAB62669
 ID AAB62669 standard; Protein; 430 AA.

XX AAB62669;

DT 23-JUL-2001 (first entry)

DE Truncated phosphatonin polypeptide (truncated MEPE).

XX Metastatic-tumour excreted phosphaturic element; MEPE; phosphatonin;
 KW phosphate; vitamin-D; skeletal formation; mineralization; truncated;
 KW osteopathic; antlyout; cytostatic; human.

OS Homo sapiens.

FH Key Location/Qualifiers

FT Misc-difference 101..192

FT /note= "the amino acid residues in this region are
 indicated incorrectly in the sequence provided
 in the sequence listing"

PN WO200132878-A2.

XX 10-MAY-2001.

XX 31-OCT-2000; 2000MO-EP10747.

XX 04-NOV-1999; 99US-0434185.
 PR 08-NOV-1999; 99GB-0026424.

XX (UNLO) UNIV COLLEGE LONDON.

XX Rowe P;

PI WPI: 2001-343487/36.

DR N-PSDB; AAF83759.

XX New phosphatonin polypeptide a regulator of phosphate metabolism, for
 PT diagnosing and treating disorders of phosphate, vitamin-D metabolism,
 PT skeletal formation e.g. osteoporosis, Paget's disease, gout

XX Claim 4: Fig 8; 135pp: English.

PS The invention relates to a novel human protein, metastatic-tumour
 CC excreted phosphaturic element (MEPE) or phosphatonin (modulator of
 CC phosphate and vitamin-D metabolism). The phosphatonin (modulator of
 CC polynucleotides and specific antibodies are useful for treating a
 CC disorder of phosphate or vitamin D metabolism, skeletal formation and
 CC mineralization. Phosphatonins are used to treat hyperphosphatemia, and
 CC osteodystrophy, secondary hyperparathyroidism, osteitis fibrosa cystica
 CC or gout. It is used to prepare a medicament for treating x-linked
 CC hypophosphatemic rickets, hereditary hypophosphatemic rickets with
 CC hypercalcaemia (HHRH), hypomineralized bone lesions, stunted growth in
 CC juveniles, oncogenic hypophosphatemic osteomalacia, renal phosphate
 CC leakage, renal osteodystrophy, osteoporosis, vitamin-D resistant rickets,
 CC end organ resistance, renal Fanconi syndrome, autosomal rickets, Paget's
 CC disease, kidney failure, renal tubular acidosis, cystic fibrosis or
 CC sprue. Phosphatonin polynucleotides are useful as molecular weight
 CC markers on Southern gels, as diagnostic probes for detecting the presence
 CC of a specific mRNA. Phosphatonin polypeptides are also useful for
 CC identifying agonists and antagonists, compounds which bind to
 CC phosphatonin and drug candidates for therapy of phosphate metabolism
 CC disorders. The present sequence represents a truncated form of
 CC phosphatonin (MEPE).

SO Sequence 430 AA:

Query Match 100.0%; Score 2279; DB 22; Length 430;
 Best Local Similarity 100.0%; Pred. No. 1.7e-181;
 Matches 430; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VNKEYSISNKNENTHNGLRMSIYPRKSTGNKGFEDDDAISKLHQEYGAALLRNMQHTM 60
 Db 1 VNKEYSISNKNENTHNGLRMSIYPRKSTGNKGFEDDDAISKLHQEYGAALLRNMQHTM 60
 QY 61 GPVTAIRKLLGEENKENTPRNVLIIPASMYAKAHSKDKKKPQRDSQAOKSPVKSSTHR 120
 Db 61 GPVTAIRKLLGEENKENTPRNVLIIPASMYAKAHSKDKKKPQRDSQAOKSPVKSSTHR 120
 QY 121 IOHNDYILKHLKSVKRIIPSPFEGSGYTDLOERGDNDISPPSGGQFPKIDPGKGEATGPD 180
 Db 121 IOHNDYILKHLKSVKRIIPSPFEGSGYTDLOERGDNDISPPSGGQFPKIDPGKGEATGPD 180
 QY 181 LEGKDIQTGFAGPSEAEETHLDTKKPGYNEIPREENGNTIGTRDETAKEADAVDSLY 240
 Db 181 LEGKDIQTGFAGPSEAEETHLDTKKPGYNEIPREENGNTIGTRDETAKEADAVDSLY 240
 QY 241 EGSNDIMGSTNFKELPGREGNRYDAGSQNAHOGKVEFHYPPAPSKERKEGSSDAAEETN 300
 Db 241 EGSNDIMGSTNFKELPGREGNRYDAGSQNAHOGKVEFHYPPAPSKERKEGSSDAAEETN 300
 QY 301 YNEIPKNGKSTPKGVDSNRNQTATLNEKORFSPKSKSGLPIPSKGLNETKNEEDSFN 360
 Db 301 YNEIPKNGKSTPKGVDSNRNQTATLNEKORFSPKSKSGLPIPSKGLNETKNEEDSFN 360
 QY 361 GPSHENIITHGRKHYVYPRNQNSTNRNKGMPQKSGMRQPHSNRFRSSRRDDSESSD 420
 Db 361 GPSHENIITHGRKHYVYPRNQNSTNRNKGMPQKSGMRQPHSNRFRSSRRDDSESSD 420
 QY 421 SSSSESDGD 430
 Db 421 SSSSESDGD 430

RESULT 3

ID AAB82922
 ID AAB82922 standard; Protein; 509 AA.

XX AAB82922;

XX 21-DEC-2001 (first entry)

DE Human osteoregulin (mature polypeptide).

XX Osteoregulin; human; bone; homeostasis; adipose; calcification;
 KW atherosclerosis; osteoporosis; osteopathic; antiatherosclerotic;
 KW therapy.

OS Homo sapiens.

PN EP1130098-A2.

XX 05-SEP-2001.

XX 27-FEB-2001; 2001EP-0301768.

XX 29-FEB-2000; 2000US-185617P.

XX 22-SEP-2000; 2000US-234500P.

XX (PFIZ) PFIZER PROD INC.

XX Brown TA, De Wet JR, Gowen LC, Hames LM;

XX WPI: 2001-604111/69.

XX N-PSDB; AAH26810.

PT Novel osteoregulin polypeptide useful for regulating bone homeostasis,
 PT adiposity and calcification of atherosclerotic plaques comprises
 PT measuring the activity of osteoregulin

XX Claim 1; Page 54-55; 90pp; English.
PS

The present sequence is that of human osteoregulin mature polypeptide, i.e., lacking an N-terminal signal sequence. Osteoregulin is a novel protein which plays a role in regulating bone homeostasis, adiposity, and the calcification of atherosclerotic plaques. 2 Splice variants of human osteoregulin were identified (see also AAB28923). The invention provides novel osteoregulin proteins, nucleic acids which encode them, vectors, antibodies, host cells which express heterologous osteoregulins, and animal cells and mammals with a targeted disruption of an osteoregulin gene. The invention also provides screening assays to identify modulators of osteoregulin activity as well as methods of treating mammals for diseases or disorders associated with osteoregulin activity. The modulators of activity may be useful in the manufacture of a medicament for, as well as for treating, a mammal in need of regulation of bone mass and/or density, adiposity, vascular flexibility, and/or atherosclerotic plaque calcification (claimed), for treating and preventing osteoporosis, and for stimulating bone repair and regeneration.

QY 241 EGSNDIMGSTNFKELPGREGNVDAGSONAHQGVDFHYPPAPSKERKEGSSDAESTN 300
 DB 336 EGSNDIMGSTNFKELPGREGNVDAGSONAHQGVDFHYPPAPSKERKEGSSDAESTN 395
 QY 301 YNEIPKNGKSTRKGVDSNRNQTATLNKORFPSKGSQGLPIPSGLDNEIKNEMDSFN 360
 DB 396 YNEIPKNGKSTRKGVDSNRNQTATLNKORFPSKGSQGLPIPSGLDNEIKNEMDSFN 455
 QY 361 GPSHENIITHGRKHYHYVPHRONNSTRNKGMPOGKSGMGRPHSNRRPSSRRDSSSED 420
 DB 456 GPSHENIITHGRKHYHYVPHRONNSTRNKGMPOGKSGMGRPHSNRRPSSRRDSSSED 515
 QY 421 SSSSSSESDGD 430
 DB 516 SSSSSSESDGD 525
 RESULT 5
 ID AAB62689 standard; Protein: 525 AA.
 XX AAB62689;
 AC AAB62689;
 XX 23-JUL-2001 (first entry)
 DE Phosphatonin polypeptide (MEPE).
 XX Metastatic-tumour excreted phosphaturic element: MEPE: phosphatonin;
 KW phosphate; vitamin-D; skeletal formation; mineralization; truncated;
 KW osteopathic; antigout; cytostatic; human.
 XX Homo sapiens.
 OS WO200132878-A2.
 PN 10-MAY-2001.
 PD 31-OCT-2000; 2000MO-EP10747.
 PF 04-NOV-1999; 99US-0034185.
 PR 08-NOV-1999; 99GB-0026424.
 XX (UNLO) UNIV COLLEGE LONDON.
 PA Rowe P;
 XX WPI: 2001-343487/36.
 DR N-PSDB: AAF83764.
 XX New phosphatonin polypeptide a regulator of phosphate metabolism, for
 PT diagnosing and treating disorders of phosphate, vitamin-D metabolism,
 PR skeletal formation e.g. osteoporosis, Paget's disease, gout
 XX
 PS Claim 4; Page 133-134; 135pp; English.
 XX The invention relates to a novel human protein, metastatic-tumour
 CC excreted phosphaturic element (MEPE) or phosphatonin (modulator of
 CC phosphate and vitamin-D metabolism). The phosphatonin polypeptides,
 CC polynucleotides and specific antibodies are useful for treating a
 CC disorder of phosphate or vitamin D metabolism, skeletal formation and
 CC mineralization. Phosphatonins are used to treat hyperphosphatemia, renal
 CC osteodystrophy, secondary hyperparathyroidism, osteitis fibrosa cystica
 CC or gout. It is used to prepare a medicament for treating X-linked
 CC hypophosphatemic rickets, hereditary hypophosphatemic rickets with
 CC hypercalcaemia (HHRH), hypomineralized bone lesions, stunted growth in
 CC juveniles, oncogenic hypophosphatemic osteomalacia, renal phosphate
 CC leakage, renal osteodystrophy, osteoporosis, vitamin-D resistant rickets,
 CC end organ resistance, renal Fanconi syndrome, autosomal rickets, Paget's
 CC disease, kidney failure, renal tubular acidosis, cystic fibrosis or
 CC sprue. Phosphatonin polynucleotides are useful as molecular weight
 CC markers on Southern gels, as diagnostic probes for detecting the presence
 CC of a specific mRNA. Phosphatonin polypeptides are also useful for
 CC identifying agonists and antagonists, compounds which bind to

CC phosphatonin and drug candidates for therapy of phosphate metabolism
 CC disorders. The present sequence represents the amino acid sequence of
 CC the entire phosphatonin (MEPE).
 XX
 SQ Sequence 525 AA:
 Query Match 99.9%; Score 2276; DB 22; Length 525;
 Best Local Similarity 99.8%; Pred. No. 3.9e-181;
 Matches 429; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 YNKEYSISNKEHTHNGILRMSIYPKSTGNKGFEDGDAISKLDHDEYGAALIRNMQHIM 60
 DB 96 LNKKEYSISNKEHTHNGILRMSIYPKSTGNKGFEDGDAISKLDHDEYGAALIRNMQHIM 155
 QY 61 GPVTAIKLLGEENKENTPRNVNLNIPASMNVAKAHAKSKDKKRPQRDSQAOKSPVSKSTHR 120
 DB 156 GPVTAIKLLGEENKENTPRNVNLNIPASMNVAKAHAKSKDKKRPQRDSQAOKSPVSKSTHR 215
 QY 121 IOHNDIYLKHLISKVKKIKISDPEGSGYTDLOEGDNDISFSDGQPFKDIPEGKGEATGPD 180
 DB 216 IOHNDIYLKHLISKVKKIKISDPEGSGYTDLOEGDNDISFSDGQPFKDIPEGKGEATGPD 275
 QY 181 LEGKDIQTGFAGPSEAEESTHDTKKPGYNEIPERENGNTIGTDETAKEADAVDSL 240
 DB 276 LEGKDIQTGFAGPSEAEESTHDTKKPGYNEIPERENGNTIGTDETAKEADAVDSL 335
 QY 241 EGSNDIMGSTNFKELPGREGNVDAGSONAHQGVDFHYPPAPSKERKEGSSDAESTN 300
 DB 336 EGSNDIMGSTNFKELPGREGNVDAGSONAHQGVDFHYPPAPSKERKEGSSDAESTN 395
 QY 301 YNEIPKNGKSTRKGVDSNRNQTATLNKORFPSKGSQGLPIPSGLDNEIKNEMDSFN 360
 DB 396 YNEIPKNGKSTRKGVDSNRNQTATLNKORFPSKGSQGLPIPSGLDNEIKNEMDSFN 455
 QY 361 GPSHENIITHGRKHYHYVPHRONNSTRNKGMPOGKSGMGRPHSNRRPSSRRDSSSED 420
 DB 456 GPSHENIITHGRKHYHYVPHRONNSTRNKGMPOGKSGMGRPHSNRRPSSRRDSSSED 515
 QY 421 SSSSSSESDGD 430
 DB 516 SSSSSSESDGD 525
 RESULT 6
 ID ABB08526
 XX ABB08526 standard; protein: 525 AA.
 AC ABB08526;
 XX 23-MAY-2002 (first entry)
 DE Human phosphatonin.
 XX Human phosphatonin; cytostatic; antidiabetic; antiinflammatory;
 KW hyperphosphatemia; arteriosclerosis; heart failure;
 KW diabetic renal disease; kidney failure; cystic fibrosis.
 XX Homo sapiens.
 OS WO200198495-A1.
 PN 27-DEC-2001.
 PD 20-JUN-2001; 2001MO-JP05263.
 PF 21-JUN-2000; 2000JP-0191088.
 PR (TAKE) TAKEDA CHEM IND LTD.
 PA Kurokawa T, Yamada T, Morimoto S;
 XX WPI: 2002-139791/18.
 DR N-PSDB: ABA99159, ABA99160.

XX Phosphatonin of human origin and DNA encoding it for diagnosis and
 PT treatment of diseases associated with disorders of phosphate
 PT metabolism, e.g., hyperphosphatemia, arteriosclerosis, heart failure,
 PT diabetic renal disease and kidney failure
 XX
 XX Claim 1; Fig 1-4; 130pp; Japanese.
 XX
 CC This invention relates to human phosphatonin which has
 CC the activity of lowering blood phosphate and increasing urinary
 CC phosphate. The proteins are cytosolic, antidiabetic and
 CC antiinflammatory in their action. Phosphatonin down-regulates
 CC sodium-dependent phosphate transport in the kidney, it down-regulates
 CC 25-hydroxy-vitamin D3-1alpha-hydroxylase in the kidney and up-regulates
 CC 25-hydroxy-vitamin D3-24-hydroxylase in the kidney. Phosphatonin is
 CC useful in the diagnosis, treatment and prevention of phosphate
 CC metabolism related diseases such as hyperphosphatemia, arteriosclerosis,
 CC heart failure, diabetic renal disease, kidney failure, acute coronary
 CC disease and cystic fibrosis. This sequence represents human
 CC phosphatonin.
 CC
 XX
 XX Sequence 525 AA:
 SQ
 Query Match 99.9%; Score 2276; DB 23; Length 525;
 Best Local Similarity 99.8%; Pred. No. 3.9e-181;
 Matches 429; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VNKEYSISNKNENTHNGLRMSIYPKSTGNKGFGEDGDAISKLHDEEYGAALIRNNMOHIM 60
 DB LNKESYISNKNENTHNGLRMSIYPKSTGNKGFGEDGDAISKLHDEEYGAALIRNNMOHIM 155
 QY 61 GPVTAIKLGEENKENTPRNVLNIIIPASMYAKASHDKKKPQDSQAOKSPVSKSTHR 120
 DB 156 GPVTAIKLGEENKENTPRNVLNIIIPASMYAKASHDKKKPQDSQAOKSPVSKSTHR 215
 QY 121 IOHNDYLLKHLISKYKKIIPDFEGSGGYTDLOERGNDISPFSGDQOPFKDIPGKEATGPD 180
 DB 216 IOHNDYLLKHLISKYKKIIPDFEGSGGYTDLOERGNDISPFSGDQOPFKDIPGKEATGPD 275
 QY 181 LEGKDIOGFAGPSEASTHLDTKKPGYNEIPEERENGNTIGTRDEYAKADAVDSLV 240
 DB 276 LEGKDIOGFAGPSEASTHLDTKKPGYNEIPEERENGNTIGTRDEYAKADAVDSLV 335
 QY 241 EGSNDIMGSTNFKELPGREGNRVDAGSONAHQKVEFYHPPAPSKERKEGSSDAEESTN 300
 DB 336 EGSNDIMGSTNFKELPGREGNRVDAGSONAHQKVEFYHPPAPSKERKEGSSDAEESTN 395
 QY 301 YNEIPKNGKSTRKGVDSHNNQATLNKQRPSPKSGKSGULPIPSRGIDNEIKNEMDSFN 360
 DB 396 YNEIPKNGKSTRKGVDSHNNQATLNKQRPSPKSGKSGULPIPSRGIDNEIKNEMDSFN 455
 QY 361 GPHENITTHGRKYHYVHRONNSTRKNGMPQKSGMWGROPHSNRFRSSRRDDSSSED 420
 DB 456 GPHENITTHGRKYHYVHRONNSTRKNGMPQKSGMWGROPHSNRFRSSRRDDSSSED 515
 QY 421 SSSSESSEDGD 430
 DB 516 SSSSESSEDGD 525

RESULT 7
 AAB82923
 ID AAB82923 standard; Protein: 540 AA.
 XX
 XX AAB82923:
 XX
 XX 21-DEC-2001 (first entry)
 XX
 DE Human osteoregulin (mature polypeptide).
 XX
 XX Osteoregulin; human; bone; homeostasis; adipose; calcification;
 KW arteriosclerosis; osteoporosis; osteopathia; arteriosclerotic;
 KW therapy.

XX
 OS Homo sapiens.
 XX
 XX EPI130098-A2.
 PN
 XX
 PD 05-SEP-2001.
 XX
 XX 27-FEB-2001; 2001EP-0301768.
 PF
 XX 29-FEB-2000; 2000US-185617P.
 PR 22-SEP-2000; 2000US-234500P.
 XX
 PA (PFIZ) PFIZER PROD INC.
 XX
 PI Brown TA, De Wet JR, Gowen LC, Hames LM;
 DR N-PSDB; AAH26811.
 DR
 XX WPI; 2001-604111/69.
 XX
 PT Novel osteoregulin polypeptide useful for regulating bone homeostasis,
 PT adiposity and calcification of atherosclerotic plaques comprises
 PT measuring the activity of osteoregulin -
 PT
 XX
 XX
 XX
 PS Claim 1; Page 59-61; 90pp; English.
 XX
 XX The present sequence is that of human osteoregulin mature
 XX polypeptide, i.e. lacking an N-terminal signal sequence.
 CC Osteoregulin is a novel protein which plays a role in regulating
 CC bone homeostasis, adiposity, and the calcification of
 CC atherosclerotic plaques. 2. Splice variants of human osteoregulin
 CC were identified (see also AAB82922). The invention provides novel
 CC osteoregulin proteins, nucleic acids which encode them, vectors, and
 CC antibodies, host cells which express heterologous osteoregulins, and
 CC animal cells and mammals with a targeted disruption of an
 CC osteoregulin gene. The invention also provides screening assays
 CC to identify modulators of osteoregulin activity as well as methods
 CC of treating mammals for diseases or disorders associated with
 CC osteoregulin activity. The modulators of activity may be useful
 CC in the manufacture of a medicament for, as well as for treating, a
 CC mammal in need of regulation of bone mass and/or density, adiposity,
 CC vascular flexibility, and/or atherosclerotic plaque calcification
 CC (claimed), for treating and preventing osteoporosis, and for
 CC stimulating bone repair and regeneration.
 CC
 XX
 SQ Sequence 540 AA:
 Query Match 99.9%; Score 2276; DB 22; Length 540;
 Best Local Similarity 99.8%; Pred. No. 4e-181;
 Matches 429; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VNKEYSISNKNENTHNGLRMSIYPKSTGNKGFGEDGDAISKLHDEEYGAALIRNNMOHIM 60
 DB LNKESYISNKNENTHNGLRMSIYPKSTGNKGFGEDGDAISKLHDEEYGAALIRNNMOHIM 170
 QY 61 GPVTAIKLGEENKENTPRNVLNIIIPASMYAKASHDKKKPQDSQAOKSPVSKSTHR 120
 DB 171 GPVTAIKLGEENKENTPRNVLNIIIPASMYAKASHDKKKPQDSQAOKSPVSKSTHR 230
 QY 121 IOHNDYLLKHLISKYKKIIPDFEGSGGYTDLOERGNDISPFSGDQOPFKDIPGKEATGPD 180
 DB 231 IOHNDYLLKHLISKYKKIIPDFEGSGGYTDLOERGNDISPFSGDQOPFKDIPGKEATGPD 290
 QY 181 LEGKDIOGFAGPSEASTHLDTKKPGYNEIPEERENGNTIGTRDEYAKADAVDSLV 240
 DB 291 LEGKDIOGFAGPSEASTHLDTKKPGYNEIPEERENGNTIGTRDEYAKADAVDSLV 350
 QY 241 EGSNDIMGSTNFKELPGREGNRVDAGSONAHQKVEFYHPPAPSKERKEGSSDAEESTN 300
 DB 351 EGSNDIMGSTNFKELPGREGNRVDAGSONAHQKVEFYHPPAPSKERKEGSSDAEESTN 410
 QY 301 YNEIPKNGKSTRKGVDSHNNQATLNKQRPSPKSGKSGULPIPSRGIDNEIKNEMDSFN 360
 DB 411 YNEIPKNGKSTRKGVDSHNNQATLNKQRPSPKSGKSGULPIPSRGIDNEIKNEMDSFN 470

QY 361 GPSHENITTHGKRYHYVPHRONNSTRNKGMPOGKSGWGRQPHSNRRSSRRDDSSSSD 420
 |||||||
 Db 471 GPSHENITTHGKRYHYVPHRONNSTRNKGMPOGKSGWGRQPHSNRRSSRRDDSSSSD 530
 |||||||
 QY 421 SSSSESDDG 430
 |||||||
 Db 531 SSSSESDDG 540
 |||||||

RESULT 8
 AAB82921
 ID AAB82921 standard; Protein; 556 AA.
 AC AAB82921;
 XX
 DT 21-DEC-2001 (first entry)
 XX
 DE Human osteoregulin.
 XX
 KW Osteoregulin; human; bone; homeostasis; adipose; calcification;
 KM atherosclerosis; osteoporosis; osteopathic; antihypertensive;
 XX therapy.
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..37
 FT Protein /label= Signal_peptide
 FT /label= Mature_protein
 FT Misc-difference 57
 FT /note= "encoded by TCA"
 XX
 XX EP130098-A2.
 PD 05-SEP-2001.
 XX
 PF 27-FEB-2001; 2001EP-0301768.
 XX
 PR 29-FEB-2000; 2000US-185617P.
 PR 22-SEP-2000; 2000US-234500P.
 XX
 PA (Pfizer) PFIZER PROD INC.
 PI Brown TA, De Wet JR, Gowen LC, Hames LM;
 DR WPI: 2001-604111/69.
 DR N-PSDB: AAB26809.
 XX
 PT Novel osteoregulin polypeptide useful for regulating bone homeostasis,
 PT adiposity and calcification of atherosclerotic plaques comprises
 FT measuring the activity of osteoregulin -
 XX
 Claim 1: Page 48-49: 90pp: English.

The present sequence is that of human osteoregulin, a novel protein which plays a role in regulating bone homeostasis, adiposity, and the calcification of atherosclerotic plaques. The sequence is predicted from the nucleotide sequence (see AAB26809) of isolated osteoblast cDNA. A splice variant of human osteoregulin was also identified (see AAB82920). The invention provides novel osteoregulin proteins, nucleic acids which encode them, vectors, antibodies, host cells which express heterologous osteoregulins, and animal cells and mammals with a targeted disruption of an osteoregulin gene. The invention also provides screening assays to identify modulators of osteoregulin activity as well as methods of treating mammals for diseases or disorders associated with osteoregulin activity. The modulators of activity may be useful in the manufacture of a medicament for, as well as for treating, a mammal in need of regulation of bone mass and/or density, adiposity, vascular flexibility, and/or atherosclerotic plaque calcification (claimed), for treating and preventing osteoporosis, and for

CC stimulating bone repair and regeneration.
 XX
 SQ Sequence 556 AA:
 Query Match 99.9%; Score 2276; DB 22; Length 556;
 Best Local Similarity 99.8%; Pred. No. 4.2e-181;
 Matches 429; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VNKEYSISNKNENTHNGIRMSIYKSTGNGKGFDDGDAISKLDHDEYGAALIRNMQHIM 60
 :|||||
 Db 127 LNKEYSISNKNENTHNGIRMSIYKSTGNGKGFDDGDAISKLDHDEYGAALIRNMQHIM 186
 |||||||
 QY 61 GPVTAIKLLGEENKENTPNNVLIIPASNNYAKAHSKDKKPPORDSOAKSPVSKSTHR 120
 |||||||
 Db 187 GPVTAIKLLGEENKENTPNNVLIIPASNNYAKAHSKDKKPPORDSOAKSPVSKSTHR 246
 |||||||
 QY 121 IOHNIDYLLKHLKSKKIPSPDFSGYTLQERGDNDISPFSDGQPFKDIQKGEATGPD 180
 |||||||
 Db 247 IOHNIDYLLKHLKSKKIPSPDFSGYTLQERGDNDISPFSDGQPFKDIQKGEATGPD 306
 |||||||
 QY 181 LEGKDIQTGFAGPSAEATHLDTKKPGYNEIPEERENGNTIGTRDETAKEDADVSLV 240
 |||||||
 Db 307 LEGKDIQTGFAGPSAEATHLDTKKPGYNEIPEERENGNTIGTRDETAKEDADVSLV 366
 |||||||
 QY 241 EGSNDIMSTNKELEPGREGNRYVDGSONAHQKVEFYPPAPSKREKKEGSSDAESTN 300
 |||||||
 Db 367 EGSNDIMSTNKELEPGREGNRYVDGSONAHQKVEFYPPAPSKREKKEGSSDAESTN 426
 |||||||
 QY 301 YNEIPKNGKSTRKGVDSNNQATLNKORFSPKSGKSGLPISRGIDNEIKNMDSEFN 360
 |||||||
 Db 427 YNEIPKNGKSTRKGVDSNNQATLNKORFSPKSGKSGLPISRGIDNEIKNMDSEFN 486
 |||||||
 QY 361 GPSHENITTHGKRYHYVPHRONNSTRNKGMPOGKSGWGRQPHSNRRSSRRDDSSSSD 420
 |||||||
 Db 487 GPSHENITTHGKRYHYVPHRONNSTRNKGMPOGKSGWGRQPHSNRRSSRRDDSSSSD 546
 |||||||
 QY 421 SSSSESDDG 430
 |||||||
 Db 547 SSSSESDDG 556
 |||||||

RESULT 9
 AAE13227
 ID AAE13227 standard; Protein; 525 AA.
 AC AAE13227;
 XX
 DT 12-FEB-2002 (first entry)
 XX
 DE Human oncogenic osteomalacia-related protein-1 (OOM-1).
 XX
 KW Human; oncogenic osteomalacia-related protein 1; OOM1; gene therapy;
 KW phosphate homeostasis-related disease; rickets; osteomalacia; cardiac;
 KW ribodysplasia; cardiomyopathy; tumoural calcinosis; renal failure;
 KW bone mineralisation; phosphaturia; cellular process.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..16
 FT /label= Signal_peptide
 FT Protein 17..525
 FT /note= "Mature oncogenic osteomalacia-related protein-1 (OOM-1)"
 FT Domain 130..142
 FT /note= "Calcium binding motif"
 FT Domain 235..258
 FT /note= "Calcium binding motif"
 FT Region 247..249
 FT /note= "Cell attachment site"
 FT Domain 264..275
 FT /note= "Calcium binding motif"
 FT Domain 412..424

FT	Modified-site	442..455	/note="Calcium binding motif"
FT	Modified-site	/note="Potential PKA phosphorylation site"	
FT	Modified-site	477..481	/note="Glycosylation site"
FT	Domain	500..503	/note="Bacterial extracellular solute-binding protein motif"
FT			
XX	MO200172826-A2.		
XX			
XX	04-OCT-2001.		
XX			
PD	22-MAR-2001; 2001WO-US09289.		
XX			
PR	24-MAR-2000; 2000US-191786P.		
PR	19-OCT-2000; 2000US-241598P.		
XX			
XX	(GEN2) GENZYME CORP.		
PA	(UYJO) UNIV JOHNS HOPKINS.		
XX			
PI	Schiavi S, Madden S, Manavalan P, Levine MDW, Jan De Beur S;		
XX	WPI: 2002-010720/01.		
XX			
PT	New polynucleotide encoding a polypeptide useful for identifying		
PT	polynucleotide expression or polypeptide activity modulators used for		
PT	treating oncogenic osteomalacia, comprises the oncogenic		
PT	osteomalacia-related gene		
PS	Claim 2; Fig 2; 65pp; English.		
XX			
CC	The invention relates to oncogenic osteomalacia-related protein 1		
CC	designated OOM1 (oncogenic osteomalacia-related factor) and its DNA		
CC	molecule. OOM1 protein is useful for treating a phosphate homeostasis-		
CC	related disease such as x-linked hypophosphataemia rickets, oncogenic		
CC	osteomalacia, rhabdomyolysis, cardiomyopathy, tumoural calcinosis and		
CC	renal failure. OOM1 proteins and DNA's are useful for modulating the		
CC	phenotype of a neoplastic cell associated with oncogenic osteomalacia;		
CC	modulating bone mineralisation; treating a disease characterised by		
CC	abnormal bone mineralisation; and treating an oncogenic osteomalacia-		
CC	associated symptom such as hypophosphataemia, phosphaturia, low serum		
CC	concentrations of 1,25-dihydroxy vitamin D or osteomalacia. OOM1 is		
CC	useful for modulating renal phosphate transport, which involves altering		
CC	the activity of OOM1 within a cell and then administering serine protease		
CC	11 to the cell. OOM1 DNA is useful for generating single nucleotide		
CC	animal models, and for searching and identifying single nucleotide		
CC	polymorphisms which are mutants, variants of the gene in human		
CC	population. OOM1 is useful as an immunogen to produce antibodies against		
CC	OOM1 and in vitro assays to screen for agents or drugs, which modulate		
CC	cellular processes. The present sequence is human oncogenic		
CC	osteomalacia-related protein-1 (OOM-1).		
CC	Note: The present sequence states that this sequence is encoded by the		
CC	DNA sequence shown in SEQ ID NO: 1 (AA021889) of the specification.		
CC	However this does not appear to be the case.		
XX			
XX	Sequence 525 AA:		
QY	Query Match	99.6%;	Score 2269; DB 23; Length 525;
QY	Best Local Similarity	99.5%;	Pctd. No. 1.5e-180;
QY	Matches 428; Conservative	1; Mismatches 1;	Indels 0; Gaps
DB	1 VNRKYSISNKEKNTNGRLMSIYPKSTGKNGFEDDDAISKLHDEVEYGAALIRNMQHIM 60		
DB	96 LNRKYSISNKEKNTNGRLMSIYPKSTGKNGFEDDDAISKLHDEVEYGAALIRNMQHIM 155		
QY	61 GPVTAIKLGEENKENTPRNVLTITIPASMTAKAHSKDKKRPORDSQAKSPYKSKSTHR 120		
DB	156 GPVTAIKLGEENKENTPRNVLTITIPASMTAKAHSKDKKRPORDSQAKSPYKSKSTHR 215		
QY	121 IOHHIDVLKHLISKYKKKIPDSFGSGYTPDJORGDNDISPFSGDGGCPFFDIPGKGATSPD 180		
DB	216 IOHHIDVLKHLISKYKKKIPDSFGSGYTPDJORGDNDISPFSGDGGCPFFDIPGKGATSPD 275		

QY	181	LEGGDIDQTGAFGSESEAE	SHLDTKKPGYNEIPERE	NGCNTIGRDEFAEADAVDSLV	240
Db	276	LEGKDIDQTGAFGSESEAE	SHLDTKKPGYNEIPERE	NGCNTIGRDEFAEADAVDSLV	335
QY	241	EGSMDIMGSTNFEELG	RGRGNRYDAGSQA	HAHGKVEFFHYPPAPSKERKEGSSDAE	300
Db	336	EGSMDIMGSTNFEELG	RGRGNRYDAGSQA	HAHGKVEFFHYPPAPSKERKEGSSDAE	395
QY	361	GPSEHNTITTGKRYHY	PHQNNSTFNKCMPOG	KGSMGOPHSNRFSSRRDDSSSSD	420
Db	456	GPSEHNTITTGKRYHY	PHQNNSTFNKCMPOG	KGSMGOPHSNRFSSRRDDSSSSD	515
QY	421	SGSSSESDDG	430		
Db	516	SGSSSESDDG	525		
RESULT	10				
AAAB82919					
ID	AAAB82919	standard; Protein: 441	AA.		
XX					
AC	AAAB82919;				
XX					
DT	21-DEC-2001	(first entry)			
XX					
DE	Mouse osteoregulin.				
XX					
KW	Osteoregulin; mouse; bone; homeostasis; adipose; calcification;				
KW	atherosclerosis; osteoporosis; osteopathic; antiatherosclerotic;				
KW	therapy.				
XX					
OS	Mus musculus.				
XX					
PN	EP1130098-A2.				
XX					
PD	05-SEP-2001.				
XX					
PF	27-FEB-2001; 2001EP-0301768.				
XX					
PK	29-FEB-2000; 2000US-185617P.				
XX					
PR	22-SEP-2000; 2000US-234500P.				
XX					
PA	(PF12) PEIZER PROD INC.				
XX					
PI	Brown TA, De Wet JR, Gowen LC, Hames LM;				
XX					
DR	WPI: 2001-60A111/69.				
XX					
PT	N-PSDB: AAH26807.				
PT					
PT	Novel osteoregulin polypeptide useful for regulating bone homeostasis				
PT	adiposity and calcification of atherosclerotic plaques comprises				
PT	measuring the activity of osteoregulin -				
XX					
PS	Claim 1; Page 43-44; 90pp; English.				
XX					
CC	The present sequence is that of mouse osteoregulin, a novel protein				
CC	which plays a role in regulating bone homeostasis, adiposity, and				
CC	the calcification of atherosclerotic plaques. The sequence is				
CC	predicted from the nucleotide sequence (see AAH26807) of a mouse				
CC	tibia cDNA clone. The invention provides osteoregulin proteins,				
CC	which express heterologous osteoregulins, and animal cells and mammals				
CC	express heterologous osteoregulins, and animal cells and mammals				
CC	express heterologous osteoregulins, and animal cells and mammals				
CC	also provides screening assays to identify modulators of				
CC	osteoregulin activity as well as methods of treating mammals for				
CC	diseases or disorders associated with osteoregulin activity. The				
CC	modulators of activity may be useful in the manufacture of a				
CC	medicament for, as well as for treating, a mammal in need of				

CC regulation of bone mass and/or density, adiposity, vascular
 CC flexibility, and/or atherosclerotic plaque calcification (claimed),
 CC for treating and preventing osteoporosis, and for stimulating bone
 CC repair and regeneration.

Sequence 441 AA;

Query Match 43.2%; Score 983.5; DB 22; Length 441;
 Best Local Similarity 50.5%; Pred. No. 1.7e-73;
 Matches 217; Conservative 52; Mismatches 134; Indels 27; Gaps 8;

QY 6 SISKENTHNLMSITPKSTGNKGFEDGDAISKLDQDEYGAALIRNMQHIMGPYA 65
 DB 34 SCGNDSITHNDLASVYPTDPTVDEGTEDGOGALLHPGQDRYGAALIRNTQVKSLEYG 93
 QY 66 IKLGEENKENTPRVNLITIPASMYAKAHSKDKKKPQRDSQAKSPVKS---KSTHRIQ 122
 DB 94 AELRREGNOEKRPQSVLVIADVNDKAVSLKIKIKQESTYLLTQSSPVKSKTKTKHTRQTR 153
 QY 123 HNIDYKHLKSKVKKISDFEGSGYTDLOERGNDISPFGDGGPFPKIDIGKRA--TGPD 181
 DB 154 RSTHYLTHLPQIKTKTSDLEGSGPDLLVKGNDVYPFGSDGQHFMHTRGKGAGSGPR- 212
 QY 182 EGDIDTGFAGPSEASTHLDTKKPGYNEIPEREENGNTICTRDETAKEDADVSLYE 241
 DB 213 --SSTRPLSGSSKAEVIDPHMSGSGSNEIPREGHGSAIYTRKAAAGAGSAGSLVG 270
 QY 242 GSDINGSTNFKELPGRGNRRDAGSONAHQKVEFHYPPAPSKERKREGSDAESTNY 301
 DB 271 GSEITGSGTNEFELPEKRNINAGSONAHQKVEFHYPPVARSRYKGVGVAHAGRA-GY 329
 QY 302 NEIPKNGSKTRKGVDSNRNOATLNEKORFSPKSGSGGLPISGLDNETIKNEMDSFNG 361
 DB 330 NEIPKSSKSSKDAEESGNOLITLTAQRFPKSGSGPALPSHLSNEVKS----- 383
 QY 362 PSHENITIGKRYHYVPRONNSTRNKGMPQKSGM--GROPHSNRFRSSRRDSSSSSD 420
 DB 384 ---EN-----HYVFHGNLTPNKMGMSQRRGSMPSRRNRNRRASTRQR--DSSESS 431
 QY 421 SSSSESDDG 430
 DB 432 SSSSESHGD 441

RESULT 11

AAB82918 standard; Protein; 435 AA.

XX AAB82918;
 AC AAB82918;
 XX 21-DEC-2001 (first entry)
 DT XX
 DE Rat osteoregulin.
 XX
 KW Osteoregulin; rat; bone; homeostasis; adipose; calcification;
 KW atherosclerosis; osteoporosis; osteopathic; antiarteriosclerotic;
 KW therapy.
 XX
 OS Rattus norvegicus.
 XX
 FH key Location/Qualifiers
 FT Peptide 1..16
 FT /label= Signal_peptide
 FT Protein 17..435
 FT /label= Mature_protein
 XX
 PN EPI130098-A2.
 PD 05-SEP-2001.
 PD 27-FEB-2001; 2001EP-0301768.
 PF 29-FEB-2000; 2000US-185617P.
 XX
 PR

PR 22-SEP-2000; 2000US-234500P.

XX (PF12) PFIZER PROD INC.

PI Brown TA, De Wet JR, Gowen LC, Hames LM;

DR WPI: 2001-604111/69.

XX N-PSDB: AAH26806.

PT Novel osteoregulin polypeptide useful for regulating bone homeostasis,
 PT adiposity and calcification of atherosclerotic plaques complexes
 PT measuring the activity of osteoregulin -

Claim 1; Page 41-42; 90pp; English.

The present sequence is that of rat osteoregulin, a novel protein
 which plays a role in regulating bone homeostasis, adiposity, and
 the calcification of atherosclerotic plaques. The sequence is
 predicted from the nucleotide sequence (see AAH26806) of a cDNA
 clone isolated from a cDNA library enriched for genes induced
 during osteoblastic differentiation. The invention provides novel
 osteoregulin proteins, nucleic acids which encode them, vectors,
 antibodies, host cells which express heterologous osteoregulins, and
 animal cells and mammals with a targeted disruption of an
 osteoregulin gene. The invention also provides screening assays
 to identify modulators of osteoregulin activity as well as methods
 of treating mammals for diseases or disorders associated with
 osteoregulin activity. The modulators of activity may be useful
 in the manufacture of a medicament for, as well as for treating, a
 mammal in need of regulation of bone mass and/or density, adiposity,
 (claimed), for treating and preventing osteoporosis, and for
 stimulating bone repair and regeneration.

Sequence 435 AA;

Query Match 41.2%; Score 938.5; DB 22; Length 435;
 Best Local Similarity 49.4%; Pred. No. 9.4e-70;
 Matches 211; Conservative 45; Mismatches 146; Indels 25; Gaps 8;

QY 9 NKEHTHGLRMSITPKSTGNKGFEDGDAISKLDQDEYGAALIRNMQHIMGPYTAIKL 68
 DB 29 NOGNIH---LASVPEPVGKCTEGGRDAPILLDONRGATLLRNITQPVKSLVTGTVE 85
 QY 69 LGSEENKENTPRVNLITIPASMYAKAHSKDKKKPQRDSQAKSPVKSSTHRIQNIIDYL 128
 DB 86 QSDRNEKPKQSVLSVLPDVHNTDYSDETEHQDRDLLQNSPQSKHTRARSTRYL 145
 QY 129 KHLKSVKRIIPDFEGSGYTDLOERGNDISPFGDGGPFPKIDIGKRA--TGPDLE---GKD 185
 DB 146 THLPQIRKILSDFEDSASPDLLVKGNDVYPFGSDGQHFMHTRGKGAGSGVSDPSSAGHP 205
 QY 186 IQTGFAGPSEASTHLDTKKPGYNEIPEREENGNTICTRDETAKEDADVSLYEVSND 245
 DB 206 V-----SGSSNVEIYDPHTNGLSGSNEIPREGHIGCAVATRGKTQAGSGADSVLSVESSNE 261
 QY 246 IMGSTNFKELPGRGNRRDAGSONAHQKVEFHYPPAPSKERKREGSDAESTNYNEIP 305
 DB 262 ITGSTNFKELPGRGNRRDAGSONAHQKVEFHYPPAPSKERKREGSDAESTNYNEIP 321
 QY 306 KNGKSTRKGVDSNRNOATLNEKORFSPKSGSGGLPISGLDNETIKNEMDSFNGPSHE 365
 DB 322 KSSKSGASKDAEESGNOLITLTAQRFPKSGSGPALPSHLSNEVKS-----SHSLGNEYKSEDSNLSLR 377
 QY 366 NI-ITHGKRYHYVPRONNSTRNKGMPQKSGM--GROPHSNRFRSSRRDSSSSSDSGS 423
 DB 378 GIALAHRRTSH-----PTRNGMSQRRGSMPSRRNRNRRASTRQR--DSSESSSSGS 428
 QY 424 SSESDDG 430
 DB 429 SSESDDG 435


```
XX  Unidentified.
OS
XX
XX  WO200214360-A1.
PN
XX
XX  21-FEB-2002.
PD
XX
XX  14-AUG-2001; 2001WO-US25542.
PF
XX
XX  16-AUG-2000; 2000US-0641034.
PR
XX  19-MAR-2001; 2001US-0812483.
PA
XX  (BIGB-) BIG BEAR BIO INC.
PI
XX  Kumagai Y, Blacher RW, Yoneda T;
DR  WPI; 2002-291971/33.
XX
XX  New peptide compound useful for reducing bone loss, is capable of
PT  enhancing bone growth, and comprises an integrin binding motif,
PR  glycosaminoglycan binding motif or a calcium binding motif -
PS  Disclosure: Page 11; 50pp; English.
XX
XX  The invention relates to a peptide compound capable of enhancing bone
CC  growth, and comprising 10-50 amino acids in a sequence, where the amino
CC  acids are in D- or L- conformation and the sequence comprises a motif
CC  selected from an integrin binding motif, a glycosaminoglycan binding
CC  motif and a calcium binding motif. The peptide of the invention is useful
CC  for reducing bone loss and for reducing renal phosphate excretion in an
CC  individual. The peptide is useful for promoting regeneration of alveolar
CC  bone and/or teeth, and increases the number and activity of odontoblasts
CC  /osteoclasts that help form dental tissues. The peptide is also useful
CC  for treating or preventing a condition associated with skeletal loss or
CC  weakness. This sequence represents a protein of a matrix extracellular
CC  phosphoglycoprotein containing an RGD motif of the invention.
XX
SQ  Sequence 97 AA;
      Query Match 22.6%; Score 514; DB 23; Length 97;
      Best Local Similarity 100.0%; Pred. No. 2.9e-35;
      Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY  105 DSOAQKSPVKSSTHRIQHNIDYKHLKSKVKKIPSPFEGSGYTDLOERGDNDISPSGDC 164
DB  1 DSOAQKSPVKSSTHRIQHNIDYKHLKSKVKKIPSPFEGSGYTDLOERGDNDISPSGDC 60
OY  165 QPFKDIPIGKGATGPDLEKGDIGTGFGAPSEASTHL 201
DB  61 QPFKDIPIGKGATGPDLEKGDIGTGFGAPSEASTHL 97
RESULT 15
AAU93682
ID  AAU93682 standard; protein; 47 AA.
XX
XX  AAU93682;
AC
XX
XX  02-JUL-2002 (first entry)
DT
XX
XX  Dental product bone growth enhancing peptide #2.
DE
XX
XX  Dental product: toothpaste; mouthwash; dental floss; bone growth;
KM  integrin binding motif; RGD; skeletal disease; dental disease; tooth;
KM  alveolar bone growth; osteoblast; odontoblast; osteopathic.
XX
XX  Synthetic.
OS
XX
XX  WO200213775-A1.
PN
XX
XX  21-FEB-2002.
PD
XX
XX  09-AUG-2001; 2001WO-US25101.
PF
```

```
XX  16-AUG-2000; 2000US-225879P.
PR
XX
XX  (BIGB-) BIG BEAR BIO INC.
PA
XX
XX  Yoneda T, Nomizu M, Kumagai Y;
DR  WPI; 2002-329525/36.
XX
XX  Dental product useful for treating skeletal diseases e.g. dental
PT  diseases comprises a base material and a compound comprising specific
PR  amino acid in a sequence containing the integrin binding motif -
PS  Claim 6; Page 19; 44pp; English.
XX
XX  The present invention relates to dental products such as toothpastes,
CC  mouthwash and dental floss comprising a base material and a compound
CC  which promotes bone growth. Such compounds are peptide sequences
CC  comprising 10-50 amino acids and containing an integrin binding
CC  motif such as RGD in the D- or L- form, preferably the L-configuration.
CC  The peptides of the invention are useful for treating or preventing
CC  skeletal diseases such as dental disease. The peptides enhance tooth
CC  and/or alveolar bone growth on areas where deterioration has occurred,
CC  as well as the growth or recruiting of osteoblast or odontoblast cells
CC  on the surface of the new skeletal growth. AAU93681-AAU93726 represent
XX  bone growth enhancing peptides for use in dental products.
XX
SQ  Sequence 47 AA;
      Query Match 10.8%; Score 246; DB 23; Length 47;
      Best Local Similarity 100.0%; Pred. No. 2.3e-13;
      Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY  108 AQKSPVKSSTHRIQHNIDYKHLKSKVKKIPSPFEGSGYTDLOERGD 154
DB  1 AQKSPVKSSTHRIQHNIDYKHLKSKVKKIPSPFEGSGYTDLOERGD 47
```

Search completed: April 11, 2003, 12:02:07
Job time : 77 secs

GenCore version 5.1.4-P5_A578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2003, 12:00:27 ; Search time 22 Seconds
(without alignments)
1878.991 Million cell updates/sec

Title: US-09-700-696B-2

Perfect score: 2279
Sequence: 1 VNKEXISXNKENTHNGLRMS.....RRDSSSSSDSGSSSESDDG 430

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 segs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : PIR_73:*

1: PIR1:*\n2: PIR2:*\n3: PIR3:*\n4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	146	6.4	667	2 A40713	cyliclin I - bovine
2	144.5	6.3	665	2 B71609	hypothetical prote
3	143.5	6.3	1207	2 T23734	hypothetical prote
4	142.5	6.3	635	2 T09648	nucleolin homolog
5	142	6.2	1445	2 T14913	CAR8 protein - yea
6	140.5	6.2	2248	2 A35938	profilaggrin - hum
7	140.5	6.2	2845	2 T45050	adenomatous polyp
8	140	6.1	772	2 JCA636	transcripton elon
9	140	6.1	852	2 T06310	hypothetical prote
10	139.5	6.1	848	2 T28055	hypothetical prote
11	138.5	6.1	1035	2 T16588	hypothetical prote
12	138	6.1	599	2 T39990	hypothetical prote
13	137	6.0	611	2 T06458	nucleolin homolog
14	137	6.0	1125	2 E90598	membrane nucleas
15	136.5	6.0	585	2 T47364	hypothetical prote
16	136.5	6.0	723	2 A48217	single-strand DNA/
17	136	6.0	1210	2 I39410	AF-4 protein, sp1
18	135.5	5.9	723	2 S33688	hypothetical prote
19	135.5	5.9	1130	2 T30251	repetin - mouse
20	135.5	5.9	1211	2 T42230	AF4 protein - mous
21	134.5	5.9	1070	2 T30848	Duffy receptor - P
22	134.5	5.9	1460	2 S48457	nucleoporin KAT7 -
23	134.5	5.9	1507	2 B47328	natural killer cel
24	133.5	5.9	1403	1 A47328	natural killer cel
25	133	5.8	771	2 T38616	hypothetical prote
26	133	5.8	938	2 T39006	related to yeast z
27	133	5.8	1213	2 A58198	serine/proline-ric
28	132.5	5.8	786	2 T33856	hypothetical prote
29	132	5.8	591	2 A45135	profilaggrin - hum

30	131.5	5.8	727	2 T29612	hypothetical prote
31	131	5.7	491	2 C97267	hypothetical prote
32	131	5.7	1307	2 T30887	16d nuclear prote
33	130.5	5.7	488	2 T46014	cyliclin II - bovin
34	130.5	5.7	530	2 JC7168	lens epithelium-de
35	130.5	5.7	799	2 S65192	hypothetical prote
36	130	5.7	578	2 S55102	psp2 protein - yea
37	130	5.7	1390	2 T14004	trfa protein - sll
38	130	5.7	2195	2 S61103	SCC1e protein - ye
39	129.5	5.7	677	1 CNHDB	chromogranin B pre
40	128.5	5.6	2481	1 D90011	EmtB protein (lmpo
41	128.5	5.6	2500	2 G71609	hypothetical prote
42	127.5	5.6	1068	2 G86452	hypothetical prote
43	127.5	5.6	1235	1 PWBVH	potassium transpor
44	127.5	5.6	1262	2 T30574	protein phosphatas
45	127	5.6	790	2 T05576	hypothetical prote

ALIGNMENTS

RESULT 1

A40713
cyliclin I - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 12-May-1994 #sequence_revision 12-May-1994 #text_change 05-Nov-1999
C:Accession: A40713; S35913
R:Hess, H.; Heid, H.; Franke, W.W.
J. Cell Biol. 122, 1043-1052, 1993
A:Title: Molecular characterization of mammalian cyliclin, a basic protein of the sper
A:Reference number: A40713; MUID:93359502; PMID:8354692
A:Accession: A40713
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-667 <HES>
A:Cross-references: GB:222779; NID:9396150; PIDN:CAA80456.1; PID:9396151
C:Keywords: cytoskeleton

Query Match 6.4%; Score 146; DB 2; Length 667;

Best Local Similarity 20.5%; Pred. No. 0.17; Matches 97; Conservative 70; Mismatches 190; Indels 116; Gaps 22;

QY	22	YKST---	GNNKGFEDGDDAIS-KLHDQEEYGAALLIRNNQHIMGPVTAIKLGEENKENT	77
DB	55	FKSSQPGCNKRRLRSEIQVTPRRD-----	KRNDELQKPAHIWIRSLKRRQS	105
QY	78	PRNVNLT---	PASNNY-----AKAKSK-----DKK-----	103
DB	106	PS--IMLVIRQASFRHPYTHITHSKAKSKKYKDKDKETALKKISKKDTGPEVDEKPK		163
QY	104	RSQNAOKSPVKS-----	KSTHRIQHNID-----YKHLISKVKKIPSPDESGGYTD	148
DB	164	RRKKADTKTPSKSHSQLSKSKSKSETNPESKDSISYIKQKKEKRSKD---SKEND		220
QY	149	LOERGNDISPFGDQGPDKDIPGKEATGPDEGKDIOTGF---AGPSEASTHLDTKK		205
DB	221	FESTSTRKYS-----	KSSKNNSDAVSETCSKSSNVGLVHLGEEDADSMEDMWL	271
QY	206	PGVNEIPEREENGNTIGTRDETAKADAVDSLVESGNDINGST-NPEKELPREGNRVD		264
DB	272	KNSQNSKRRPKTKD--AKDKAGKGSDAESYDSKAKDKGATKDTKKAKKIDESTD		329
QY	265	A--GQNAHOGKVEFHPYPAPEKEREKSSDAESTYNEIPKNGKSTRGVDHSNR		321
DB	330	AESGDSKAKKKE-----	SKDKKAKKADDAADASGD-SKDAKKDSKKKDKSKD	382
QY	322	NOATLNEOKFPB---	KGKSGCLPIPSRGLDNEIKNEMDSFENGSHENITTHGRKYHYV	377
DB	383	DNKKKAKKDAESTDAESDPSKAKKDKSK-----	KGKDSKDKDKKXPD---AKKDAES	432
QY	378	PHQNNSTNRKMGPGKSGWGQRPHSNRRPSSRRDSSSSSDSGSSSDG		430
DB	433	TDAESGDSKAKKDKKG-----	KKDKKAKKADAVSTDADSESGD	475

RESULT 2

B71609

hypothetical protein PF0680w - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum

C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000

C:Accession: B71609

R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;

; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.

Science 282, 1126-1132, 1998

A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.

A:Reference number: A71600; MUID:99021743; PMID:9804551

A:Accession: B71609

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-665 <GAR>

A:Cross-references: GB:AE001410; GB:AE001362; NID:g3845245; PIDN:AACT1925.1; PID:g384524

A:Experimental source: clone 3D7

C:Genetics:

A:Gene: PF0680w

Query Match

Best Local Similarity 19.6%; Score 144.5; DB 2; Length 665;
Matches 95; Conservative 72; Mismatches 197; Indels 121; Gaps 21;

QY 2 NKEYSISN-KE---NTNGLIMSIVPKSTGNKGFEDGDAISKLDQEEYGAALIRN- 54

DB 102 NKDSINNITKEKDDNNNNNGTK-QIEKKNINK-----SDLHQNLNLQSGKNE 151

QY 55 -----NQOIMGPVTAIKLLGEENKENTPRNVLNITPASNMYAKAHSKD 98

DB 152 QDINKNEKGKODISNSNAENKRDYEGYKELEKKEE-----KISDHDVE 198

QY 99 KKKPORDSOAQSPYKASTRIQHNIDYLKH-LSKYVKI-----PDDESG 145

DB 199 ENKSDHAKVEEN--KSDDHAKVEENKSDHAKIEYAKVEHEDEDEDEKKEKSEKN 256

QY 146 YTDLOERGDNDISFSGDGPFKDI-PKGATGPDLGKDIQTGFAGPSAESTHLDTK 204

DB 257 KDEKNDENDENDISDEVDVDEEDKKNENDIDDDKK-----ETDKTHLEE 306

QY 205 KPGNEIPERE-----ENGWTITGRDETAKAADVSLVGSNDI-MGSTNKTELPGR 258

DB 307 E---NEITEKEFSKKNKNGKRR-DTKREKSKDTEK-----EKSDIEKEKSKREKES 356

QY 259 EGNVVDAGSONAHQGYEFHYPPAPSKERKREGSSDA-----ESTYNEIPIKN 307

DB 357 KDKREKKGKDEKSKDIKEKEDKDIKEKSKDTAKEREKDIKEKSKMEKLN 416

QY 308 GKSTRGVDHSHRNQATLNEKORFPKSGKSQGLPIPSRLDNEIKNEMDSF--NGPSH 364

DB 417 KQNDKKKDD---NEKKKNDKDIHD-----DNDENDMEIEENDEED 458

QY 365 ENITTHRKHYHYVHRQNNSTRNKGMPQKSGWGRPHSN--RRSSRRRDDSSSSSGS 423

DB 459 EDEMEKKKKKKKKKNGENGENGENGENGENGENGENENENENENENENEN 518

QY 424 SSES 428

DB 519 ENENE 523

RESULT 3

T23754

hypothetical protein T05C12.10 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999

C:Accession: T23754; T24513

R:Thomas, K.

submitted to the EMBL Data Library, June 1995

A:Reference number: T19793

A:Accession: T23754

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1207 <MTL>

A:Cross-references: EMBL:Z49968; PIDN:CAA90265.1; GSPDB:GN00020; CESP:T05C12.10

A:Experimental source: clone M110

R:Barton, J.

submitted to the EMBL Data Library, October 1995

A:Reference number: T19901

A:Accession: T24513

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1207 <MT2>

A:Cross-references: EMBL:Z66500; PIDN:CAA91313.1; GSPDB:GN00020; CESP:T05C12.10

A:Experimental source: clone T05C12

C:Genetics:

A:Gene: CESP:T05C12.10

A:Map position: 2

A:Introns: 31/3; 87/2; 141/3; 180/2; 203/3; 267/1; 776/2; 794/2; 834/2; 1086/3; 1143/

Query Match

Best Local Similarity 25.7%; Score 143.5; DB 2; Length 1207;
Matches 81; Conservative 29; Mismatches 140; Indels 65; Gaps 13;

QY 123 HNIDYKHLKSKYKIPSPDESGYTDLOERGDNDISFSGDGPFKDIPKGEATGPDLE 182

DB 510 HSPDDNLLERKDEGNPKGKAGAG---NGNCGDQ-KNNNGKNGTGGDGGNGNGNGLT 564

QY 183 GKDIOTGFAGSSEASTHLD---TKRPYNEIPIEREENG-GNTITGRDETAKADAADV 237

DB 565 GDGNGTGDGDNNEGNGNGSDKNSGAGTKEENEGGNGNGTGDGNGDNDNGNG 624

QY 238 S--LVESNDIMGSTNFKELPREGNRYD-AGSONAHQGYEFHYPPAPSKERKREGSSD 294

DB 625 SKGLGTSGGDKGKNGNSGTPKSDGKEDGAGSGSGNGK-----EDGKKS GSG 675

QY 295 AAESTYNEIPIKNKSGSTRGVDSHNRNQAATLNEKORFPKSGKSQGLPIPSRLDNEIKN 354

DB 676 KCGAGN---KSGSDGS---GDGKNNNGNGTGDGDKR-NCKGSGSG-----DNDKSG 719

QY 355 EMDSFNPSHNITTHRKHYHYVHRQNNSTRNKGMPQKSGWGRPHSNRRSSRRRDD 414

DB 720 TRAAGKNAEGNGKNG---NDGKSGSGSDGSGAGOK----- 753

QY 415 SSES DSGSSSES DSG 429

DB 754 -GOKSDSESGENDAG 767

RESULT 4

T09648

nucleolin homolog num1 - alfalfa

N:Alternate names: num1 protein

C:Species: Medicago sativa (alfalfa)

C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000

C:Accession: T09648

R:Boegre, L.; Jonak, C.; Mink, M.; Meskiane, I.; Traas, J.; Ha, D.T.C.; Swoboda, I.;

Plant Cell 8, 417-428, 1996

A:Title: Developmental and cell cycle regulation of Alfalfa num1 a plant homolog of

A:Reference number: Z16796; MUID:96361876; PMID:8721748

A:Accession: T09648

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-635 <BOE>

A:Cross-references: EMBL:X88445; NID:g1279562; PIDN:CAA61298.1; PID:g1279563

C:Genetics:

A:Gene: num1

A:Superfamily: nucleolin; ribonucleoprotein repeat homology

Query Match

Best Local Similarity 20.2%; Score 142.5; DB 2; Length 635;
Matches 71; Conservative 48; Mismatches 137; Indels 95; Gaps 11;

QY 97 KDKKKPORDSOAQSPYKASTRIQHNIDYLKHLKSKYKIPSPDESGYTDLOERGDND 156

Db 23 KSGKGRQAEKEEVKAVSAKK-QKVEEVAKKOKALKVYKKEESSESS-----SEESD 75
QY 157 ISPPSSGCGCFKDPGCG-----EATGPDLEGKDIQGFAGPSPAEETHLDTKKPGY 208
Db 76 EKPAKPAKPAKSKTPPANGVNKAQAPPTTSESSSDI-----SSDEEVKPPAA 125
QY 209 NEIPEREENGNTIGTRDETAKEDADVSLVEGSNDIMGSTNFKELPGREGNVDAGSQ 268
Db 126 KAVPSK--NSAPYKKDDESEEE-----DSDESSSSSDEEDKPAKAVPSKNGS----- 172
QY 269 NANOGEVEHYPPAPSKERKKEGSSDAESTNNE-----IPKNGKSTRKGVNHS 319
Db 173 -----APAK---KDDSEEDSEDESDDEKPAKAVSKNGSVSAKKDDSSS 216
QY 320 NRNAOTLNEKORPSSKSGGLPIPSRGLNETIKNEMDSFNGPSHEHIIHGKRYHYVPH 379
Db 217 DEOSESEDEEDKPAKAAKSNVSAFTKKAASSDDESD----- 254
QY 380 RQNNSTRNKKMPQCKGSGWGRQPHSNRRFRSSRRDSSSSSSSSSSD 430
Db 255 -EESDEDEDAKPYSKPA-----AVAKSKKSDSDSDDEDDSDSDSD 295

RESULT 5

T14913
C:AT8 protein - yeast (Kluyveromyces marxianus var. lactis)
C:Species: Kluyveromyces marxianus var. lactis, Candida spherica
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #ext_change 09-Jun-2000
C:Accession: T14913
R:Georits, I.; Breunig, K.D.; Vandenhaute, J.
submitted to the EMBL Data Library, June 1998
A:Description: Sequence of K1CAT8.
A:Reference number: 218263
A:Accession: T14913
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1445 <GEO>
A:Cross-references: EMBL:AF070974; NID:g3228690; PID:g3228691; PIDN:AAC23607.1
C:Gene: CAT8
C:Superfamily: unassigned GAL4-type zinc cluster proteins; GAL4 zinc binuclear cluster
F:178-215/Domain: GAL4 zinc binuclear cluster homology <GL4>

Query Match 6.2%; Score 142; DB 2; Length 1445;
Best Local Similarity 21.4%; Pred. No. 0.76;
Matches 99; Conservative 48; Mismatches 150; Indels 166; Gaps 20;

QY 2 NKEYSISNKENNTINGLRLMSIYPRKSTGKGFEDDDAISKLDQEEYGALLIRNMQHMG 61
Db 1105 NNNNNNTNTNTNNNNSSRNTHNLGA---GIEDIDHGS-----LRSDASLNL 1149
QY 62 PYVAIKLGG---DENKENTP-RNVNLNIPASMYAKAHSKDKKKPQRDSQAOKSPYKSKS 117
Db 1150 LSNALNGHGSABGMDGTSVNLNLN-----STSDS 1181
QY 118 THRIQHNIDYLK-----HLKVKKIPSEDESGSYTDLQERGDNDISPFSG 162
Db 1182 LFKVPSNGDFLKDYYINNNSTGLSLNHLHTSVKGPS-----LSQGLSIFMSSGSG 1232
QY 163 DGGPFDIPKGE-ATGPDLEGKDIQGFAGPSPAEETHL-----DTKKPGYNET 211
Db 1233 TNLPSYVNRGESKHPGSVNLNKRQAPAMDGFEFADASGLAPLAWSPDK----- 1285
QY 212 PEREENGNTIGTRDETAKEDADVSLVEGSNDIMGSTNFKELPGREGNVDAGSQNAH 271
Db 1286 POLNSNNGND-GLNPATG-----TVLESQDND---STN-----ASVVOLO 1321
QY 272 QGVVEHYPPAPSKERKKEGSSDAESTNNEIPKNGKSTR-KGYDHSNRNQTILN--E 328
Db 1322 OHQOQTHNRP-----SHNSSITPMGKOSTSDRQNSHHQDNNVGSFHSMA 1365
QY 329 KQFPFKSGKSGGLPIPSRGLNETIKNEMDSFNGPSHEHIIHGKRYHYVPHRONSTRNK 388

Db 1366 SPIPEDTSMQGL-----HHOHMLGMPNSTLINDTNT 1397
QY 389 GMPQCKGSGWGRQPHSNRRFRSSRRDSSSSD---SGSSSESD 428
Db 1398 G---GANVNTNISISTNRGPRRRMNNNAASSSDPNSAGDSSVSD 1437

RESULT 6

A35938
C:Profilelagrin - human (fragments)
C:Species: Homo sapiens (man)
C:Date: 14-Dec-1990 #sequence_revision 02-Jul-1996 #ext_change 29-Sep-1999
C:Accession: A35938
R:gan, S.O.; McBride, O.W.; Idler, W.W.; Markova, N.; Steinhert, P.M.
Biochemistry 29, 9432-9440, 1990
A:Title: Organization, structure, and polymorphisms of the human profilaggrin gene.
A:Reference number: A35938; MUID:91064347; PMID:2248957
A:Accession: A35938
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-2248 <GAN>
A:Cross-references: GB:J02929
C:Gene: GDB:FLG
A:Map position: 1q21-1q21
C:Superfamily: unassigned calmodulin-related proteins; calmodulin repeat homology
C:Keywords: EF hand; epidermis; polymorphism; tandem repeat
F:246-569/Region: flaggrin repeat
F:570-893/Region: flaggrin repeat
F:1074-1397/Region: flaggrin repeat
F:1573-1896/Region: flaggrin repeat

Query Match 6.2%; Score 140.5; DB 2; Length 2248;
Best Local Similarity 19.0%; Pred. No. 1.6;
Matches 90; Conservative 70; Mismatches 213; Indels 101; Gaps 17;

QY 8 SNEKENTHNGLRMSIYPRKSTGKGFEDDDAISKLDQEEYGALLIRNMQHMGPTAIK 67
Db 320 SGGQSHST---TTQGRSDASKSSGSRSTRRTDQEOSGDSRHSNGHDEASTRAES 375
QY 68 L-----LGEENKENTPRNVNLNIPASMYAKAHSKDKKKPQRDSQAOKSPYKSKST--- 118
Db 376 SRHSQVGGGSSGSRSTR-----NQGSSFSQD-----RDSQAQSEDSRRRSASAS 420
QY 119 -----HRIQHNIDYLKHLKVKKIPSEDESGSYTDLQERGDNDISPFSG 162
Db 421 RNRGSAQEOSRDSRHPGSHOEDRAGHROSAES---SSQGTNHAENSSGCG---AAS 473
QY 163 DGGPFDIPKGEATGPDLEGKDIQGFAGPSPAEETHLDTKKPGY--NEIPEREENG 220
Db 474 SOEOANSSAGEHRSQHQSADSSRHSIGIHGQASSAVDSCHRGVSGSASDQEGHSEN 533
QY 221 TIGTRDETAKEDADVSLVEGSNDIMGSTNFKELPGREGNVDAGS---QNAHOGKYE 276
Db 534 SDOGSVSGQRQARSHQSHOESTRG-----QSGRGR---SSSFLYGVSTHQSS 582
QY 277 FHPPAPSKERKKEGSSDAESTNNEIPKNGK-----GSTRK-----VD--- 317
Db 583 AHRGSAPTSTRRROGSHHDDARDSRHSASQEGDITGRHGPSRGRGQSHYEQSVDRAG 642
QY 318 HSNRNQATLNEKORPSSKSGGLPIPSRGLNETIKNEMDSFNGPSHEHIIHGKRYHYV 377
Db 643 HSGSHSHSTTSGQSDASGISTG---SRASQRTDEQSDGTRHSG-----SHOE 692
QY 378 PHRONSTRNKKMPQCKGSGWGRQPHSNRRFRSSRRDSSSSD---SGSSSES 427
Db 693 ASYQADSSRHSQVGGQSGASGRSTRNNGSSVQDRNBOEQSDSEHSHSSASARN 746

RESULT 7

I49505

adenomatous polyposis coli protein - mouse

N:Alternate names: APC

C:Species: Mus musculus (house mouse)

C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 13-Aug-1999

C:Accession: I49505

R:Su, L.

Science: 256, 668-670, 1992

A:Title: Multiple intestinal neoplasia caused by a mutation in the murine homolog of the

A:Reference number: I49505; MUID:92263101; PMID:1350108

A:Accession: I49505

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2845 <RES>

A:Cross-references: GB:M8127; NID:g191991; PIDN:AB59632.1; PID:g191992

C:Superfamily: adenomatous polyposis coli protein

Query Match 6.2%; Score 140.5; DB 2; Length 2845;
Best Local Similarity 21.3%; Pred. No. 2.1;
Matches 108; Conservative 67; Mismatches 212; Indels 121; Gaps 23;

25 STGNKGF-----EDGDAISKLDQEEYGAALR--NNMOHMGVTAIKL----- 68
2092 SPGSNFPWKAIOEGANSIVSLHQAAAAACLSQASSDSDSLKSGISLSPFHLT 2151
69 LGEEKENTPRNVLTIPASMYAKAHSKDKKPPQRS--QAOKSPYKSTHRIQNI 125
2152 PDQEKKPTTS---NKGRILKPGKSTLEAKTIESEKKGKIGKKYKSLITGKIRNS 2207
126 DYKHLKSK--YKKIPSDGSGYTDLO--ERGDNDISPSGDPGPKDIPKGEATGPD 181
2208 EISSQMKQPLTNMPSISRGRTMIHPIGLRNSSSTSPVSKGPKPLTPAKSPSEGCA 2267
182 ECKDQTFGAPRSE--AESTHLDTKKPGYNEIPEREENGNTIGTDEFAKEADAVD--- 237
2268 TTSPTGTPAGSELSPTTROTSGISGNK-----GSSRSGSDSPSTPSRPTQPLSRP 2320
238 -----SLVEGSDNDIMSTNFKELPGREGNVDAGSONAHGKVEFYPPAPSKERKE 290
231 MGPSPRNISIPGRNISIPRNKISQLP--RTSSPSTASTKSSGKNSY---TSPGQLSQ 2376
291 G-SSDAESTNNEIPK-----NGKSTRK-----GVDSNR-- 321
2377 NLTKQASLSKNASSIPRESEASKGLNOMNGSNKKVELSRMSTKSSGESDSEPPA 2436
322 --NOAT-----LNKQRFPSKKGKQGLPIPSRGDNE--TKNEMDSFNGPSH 364
2437 LVKQSTFIKEAPSPILKRLKLESASFESLSPSSRPDSPTRSQAOTPVLSPIPDMSLSTH 2496
365 ENITTHG-RKYHYVPHRONNSTRNKGPQGR-----GSGWGRQ--PH 402
2497 PSVQAGWRK--LPENLSPTEIYNDGRPTKRHDIAKSHSESPSLPINRAGTWKREHSKH 2554
403 SNR--RSSRRRDSSESDDSSSESSED 428
2555 SSSLPRVSTWRTGSSSSITLSASSESE 2582

RESULT 8

transcription elongation factor B chain 3, elongin A - human

N:Alternate names: Ela protein; elongin complex A chain; SITI 110K subunit

C:Species: Homo sapiens (man)

C:Date: 10-May-1996 #sequence_revision 19-Jul-1996 #text_change 05-Nov-1999

C:Accession: J04636

R:ASO, T.; Haque, D.; Fukudome, K.; Brower, C.S.; Conaway, J.W.; Conaway, R.C.

Gene 168, 277-278, 1996

A:Title: A human cDNA encoding the 110-kDa A subunit of RNA polymerase II transcription

A:Reference number: J04636; MUID:96194914; PMID:8654961

A:Accession: J04636

A:Molecule type: mRNA

A:Residues: 1-772 <ASO>

A:Cross-references: GB:I47345; NID:g992562; PIDN:AAA75492.1; PID:g992563

A:Experimental source: liver

C:Comment: This factor is a heterotrimer, and is a target for transcriptional regulat

C:Genetics:

A:Gene: GDB:TCRB3; SITI; Ela

A:Cross-references: GDB:636397; OMIM:600786

A:Map position: 1p36.1-1p36.1

C:Keywords: heterotrimer; liver; transcription factor

Query Match 6.1%; Score 140; DB 2; Length 772;
Best Local Similarity 20.5%; Pred. No. 0.47; Mismatches 144; Indels 144; Gaps 17;
Matches 87; Conservative 49;

23 PKSTGNKGFEDGDAISKLDQEEYGAALIRNNMOHMGVTAIKLGEENKENTPRNVLT 82
210 PKGHSNAPFQRLGASQGRHIGEPHKGIVSQNKEH----- 245
83 NIIPASMYAKAHSKDKKPPORDSOAQSPYKSTHRIQNIIDYKLSKVKKIPSPFE 142
246 -----KSSHDKKPPVDAKSDKASVYSREKSH-----KALSK----- 277
143 GSGYTDLERGDNDISPSGDPGPKDIPKGEATGPDEKQDQTFGAPGSEAES--THL 201
278 -----EENRRPSPGNAREKP--PSSGVKKEKREBSSSLKKCLPSEASDNHL 325
202 DTKKPGYNEIPEREENGNTIGTDEFAKEADAVDSLVEGSDNDIM-----GSTNFKEL 255
326 --KKPKHND--PEKAK-----LDKSKQGLDSFDTG--KGAGDLLPKVKEKGSNNLTKP 372
256 PEREGNRYDAGSONAHGKRV-----EFHYV-----PAPSKRKRREGSSDAES 298
373 EKVYTNLDRKSLGS--LPKVEETMDEDEFQPTMSFESYLSYDOPRKRRKIVKT--SATA 430
299 TNYNEIPKNGKSTKRGVDSNRNOATLNEKORFSPSKSGOGL----- 341
431 LGDKLKKNDKSKTCKNLDVQKLPKVKTKTSEKAKAGADLAKIRKVPDVLPLDPLPA 490
342 -----PIPSRGDNEIKNEMDSFNGPSHEN-----IITGKRYHYV---- 378
491 IQANRYPLPSLILSSFPQKRAKAPSSQDEEAGFTGRMRMSKMOVYSGSKAYLPMKMT 550
379 --HRQ 381
551 LHQQ 554

RESULT 9

T06310

hypothetical protein F11C18.80 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 22-Oct-1999

C:Accession: T06310

R:Bevan, M.; Terry, N.; Ardielles, W.; Buysshaert, C.; Dasseville, R.; De Clerck, R.;

ewes, H.W.; Mayer, K.F.X.; Scheller, C.

submitted to the Protein Sequence Database, April 1999

A:Reference number: Z15589

A:Accession: T06310

A:Molecule type: DNA

A:Residues: 1-852 <BEV>

A:Cross-references: EMBL:AL049607; GSPDB:GN00062; ATSP:F11C18.80

C:Genetics:

A:Gene: ATSP:F11C18.80

A:Map position: 4

A:introns: 34/3; 102/3; 163/2; 205/3; 269/3; 282/3; 564/3; 602/3; 622/2; 639/3; 667/3

Query Match 6.1%; Score 140; DB 2; Length 852;
Best Local Similarity 21.8%; Pred. No. 0.53; Mismatches 154; Indels 86; Gaps 16;
Matches 83; Conservative 57;

67 KLIGEEKENTPRNVLTIPASMYAKAHSKDKKPPORDSOAQSPYKSTHRIQNIID 126
524 KVVSGDNNAQSESTK-----PKREKKKPGKGAIDESLHTSSGDNEKPAVS 568
127 YLKLHLSKYKKIPSDGSGYTDLERGDNDISPSGDPGPKDIPKGEATGPDELEADI 186

[illegible]

RESULT 10
T28055
hypothetical protein ZK858.1 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #ext_change 29-Oct-1999
C:Accession: T28055
R:White, S.
submitted to the EMBL Data Library, September 1996
A:Reference number: Z20462
A:Accession: T28055
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-848 <WIL>
A:Cross-references: EMBL:Z79759; PIDN:CAB02138.1; GSPDB:GN00019; CESP:ZK858.1
A:Experimental source: clone ZK858
C:Genetics:
A:Gene: CESP:ZK858.1
A:Map position: 1
A:Introns: 26/3; 82/3; 130/3; 195/3; 241/3; 613/1; 762/1

```

D6  657  -SDGSD:INVSMMNNITDQGRASSNPSRPOQTINTRNG-----PTNNIPYD  704
QY    409  SRRRDDSESSDSSGSSSESD  428
D6    705  SFRSQNKNSITLDGSSNNSSEE  724

```

```

RESULT 11
T16588
hypothetical protein K09C4.10 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 04-Mar-2000
C:Accession: T16588
R:Du, Z.
submitted to the EMBL Data Library, December 1995
A:Description: The sequence of C. elegans cosmid K09C4.
A:Reference number: Z18542
A:Accession: T16588
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1035 <DUZ>
C:Cross-references: EMBL:U43375; NID:g1125841; PID:g1125846; PIDN:AAA83622.1; CESP:K0
C:Genetics:
A:Gene: CESP:K09C4.10
A:Introns: 19/3: 65/1; 124/2: 151/3; 197/2: 263/1; 322/2: 354/2; 751/2: 813/2; 866/1;
Superfamily: Caenorhabditis elegans hypothetical protein K09C4.10

```

Db EVAAQ-----PARESKRTQTSEDKMDREFSFTFIEKFHFNQSYYP-PS----RYA 628
 ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
OY 292 SSDAASTWNWNEPKNGKSGSTRKYDHSRRNATLNEQRPPSKSGGLPIPSRGIDNE 351
 ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db 629 AGTAADS-----HKRNAHP--QRORPSIRNLISG----- 656
 ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
OY 352 IKRNMDSFGNSHENIITHGRKYHY---VPHRNQNSTNNKKMPQCKSGMGNOPSHNRRES 408
 ::: ::: ::: ::: ::: ::: ::: ::: ::: :::

RESULT 12
TJ3990
hypothetical protein SPBC25D12.02c - fission yeast (*Schizosaccharomyces pombe*)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: TJ3990
R.Lyne, M., Rajendream, M.A.; Barrell, B.G.; Whithead, S.: Chillingworth, T.: Church
submitted to the EMBL Data Library, July 1998

```

Query Match          6.0%; Score 137; DB 2; Length 1125;
Best Local Similarity 20.8%; Pred. No. 1.1;
Matches      75; Conservative      56; Mismatches 136; Indels   94; Gaps

QY  24 KSTGKNGFEDGDGDAISKLHDEEYGAALIRNNMCHIMPPVAIKLLEGENKENTPRPVYLN 83
DB  467 KDDQKNKSNKCNKNSDN-----DSKDLKPCT-----NNEQNQNTDDSK 507
QY  84 IIPASMYNAK-AHSKDKKKPQRDSQAQKPYVSKRSTHRIQHNIDYLKLHLSVKK-----I 137
DB  508 IDSASQNNSTNTNEKQDKIDSQD-ESKNNAIKSQNDQKDSNLSSAKNDTQPSKSSPOI 566
QY  138 PDSDEGSGTLDQEGDNDISPFSSDGPFPKDIPEKGATGPDLDEGKQIQGFGAPSPAE 197
DB  567 NENLENDQETSHSNGGEENDSK-----DQNTSNSQTKNDLRSQKQ----- 600
QY  198 STHLDTKKKGVY----ETPEREENGNTIGTRDE--TAKADADVSL--VEGSNDIMG 246

```

Db 609 --NLTKNPSSNSVETKNETONNENSTKKEIDTSKTOQDSTNSLNKNEKTNQVET 666
QY 249 STNFKELPREGNVDAGSONAHQGVFHHPPAPSKREKREGSSDAESTNYNEIPKNG 308
Db 667 KTNTE-----SNSN-----STNKOENSTKKEEISKSESNNVNN 701
QY 309 KGSTRK---GVDHS-----NRQATLNKORFPSPKSGKSGQLPIPSRGLDNEIKN 354
Db 702 SNSTKKEENIDNKKEEISKSESNNVNSNTNTQNETPETNESQNNVITICKNPNQSLN 761
QY 355 E 355
Db 762 Q 762

RESULT 15

T47364

hypothetical protein F7M19.70 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000

C:Accession: T47364

R:Nakamura, G.; Fartmann, B.; Dauner, D.; Sterr, W.; Holland, R.; Weichselgartner, M.;

Mayer, K.F.X.

Submitted to the Protein Sequence Database, April 2000

A:Reference number: 224458

A:Accession: T47364

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-585 <NYA>

A:Cross-references: EMBL:AL38643

A:Experimental source: cultivar Columbia; BAC clone F7M19

C:Genetics:

A:Map position: 3

A:Introns: 101/3: 233/2: 328/3: 361/2

A:Note: F7M19.70

Query Match

Best local similarity 6.0%; Score 136.5; DB 2; Length 585;

Matches 80; Conservative 42; Mismatches 164; Indels 89; Gaps 13;

QY 16 GLRMSIYPKSTGNKGFEDDDAISKLDQEEYGA-----ALIRNN 55
Db 216 GFDCIYIPKYNLSESIKAVREKLHSGEGEKKQMALEFVSTVITPKKKEGFIIDAF 275
QY 56 MOHMGPTVAIKL--GEENKENTPRNVLN-----TIP----- 86
Db 276 IYRIYDDLDACETFPWGRYTFEDNTKNIFFHMKITFKGRVQOTWCPGFLIPLKIFWREL 335
QY 87 -----ASMYAKAHSKDKKKPQDSQAQSPVSKSTHRIQHNIDYL-KHLSTVKKIPSD 140
Db 336 YDLDVASRGANTTEQQRDAFOEDQAMNAGFK-ELEKRIKRPDLCEHYKDHRCVNN 394
QY 141 FEGSGYTDLOERGDNDIPFSGDQPFKIDIGKEATG---PDEGKDITQTFAGPSEA 196
Db 395 AMNGGLNFQASDDNRGCGFQKLEADLDDEMGVG-ATGKRASDDIG---SGFGNEQE 449
QY 197 ESTHLDTKKPGYNETPEREENGNTIGTRDETAKADAVDVSIVSGSNDIMGSTNEKELP 256
Db 450 EDAS-----DEGAKEEENDDEMGVEEDAQVATDEGAEGDND-----EMA 493
QY 257 GRE---GNRVDAGSONAHQGVFHHPP-----APSKERKREGSSDAESTNYNEIPKNG 308
Db 494 GVEETSEKVDATDEGAEEASNRGPEEGDDASNREGVEEGDEMAKEKEVAEISGK 553
QY 309 KGSTRKGVDSNRNQ 323
Db 554 KTSPEKRRSRHPSQ 568

Search completed: April 11, 2003, 12:03:45
Job time: 29 secs

GenCore version 5.1.4-P5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2003, 11:57:01 ; Search time 14 Seconds

(without alignments)
1273.916 Million cell updates/sec

Title: US-09-700-696B-2

Perfect score: 2279

Sequence: 1 VNKEYSISNKNENTHNGLRMS.....RRDSSSSSDSGSSSESDCD 430

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	157	6.9	687	1	DSPP_RAT
2	152	6.7	1253	1	DSPP_HUMAN
3	151	6.6	510	1	DMP1_BOVIN
4	147.5	6.5	934	1	DSPP_MOUSE
5	146	6.4	667	1	CYLL_BOVIN
6	141.5	6.2	513	1	DMP1_HUMAN
7	141	6.2	633	1	MLH_TETRA
8	140.5	6.2	2845	1	APC_MOUSE
9	138	6.1	1181	1	NKX1_RAT
10	136	6.0	389	1	SER1_BOMMO
11	136	6.0	1210	1	AF4_HUMAN
12	135.5	5.9	722	1	SSRP_DROME
13	135.5	5.9	1130	1	REPT_MOUSE
14	134.5	5.9	503	1	DMP1_MOUSE
15	134.5	5.9	1070	1	PVDR_PLAYS
16	134.5	5.9	1453	1	NKCR_MOUSE
17	134.5	5.9	1460	1	N159_YEAST
18	133.5	5.9	1462	1	NKCR_HUMAN
19	131.5	5.8	2476	1	S3BI_XENLA
20	131	5.7	1307	1	S3BI_MOUSE
21	130.5	5.7	488	1	CYLL_BOVIN
22	130	5.7	578	1	PSP2_YEAST
23	130	5.7	1089	1	Y553_HUMAN
24	130	5.7	2195	1	SC16_YEAST
25	129.5	5.7	677	1	SCL_HUMAN
26	129.5	5.7	1311	1	ATRX_DROME
27	127.5	5.6	1235	1	TRK1_YEAST
28	127	5.6	718	1	YJ70_CORGL
29	126.5	5.6	1189	1	YJH6_YEAST
30	126	5.5	425	1	YNN6_YEAST
31	126	5.5	1403	1	YDF3_SCHPO
32	124.5	5.5	658	1	STC1_STRAU
33	124.5	5.5	843	1	CYPL_BRUMA

ALIGNMENTS

RESULT 1	ID	DSPP_RAT	STANDARD:	PRT:	687 AA.
AC	062598	09R057	P70578		
DT	15-JUN-2002	(Rel. 41, Created)			
DT	15-JUN-2002	(Rel. 41, Last sequence update)			
DE	15-JUN-2002	(Rel. 41, Last annotation update)			
DE	Dentin sialoprophosphoprotein precursor (Contains: Dentin phosphoprotein (Dent. phosphoprotein) (DPP), Dentin sialoprotein (DSP)).				
GN	DSPP OR RDP2				
OS	Rattus norvegicus (Rat).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
OX	NCBI_TaxID=10116;				
RN	[1]				
RP	SEQUENCE FROM N.A. (ISOFORM DPP-2).				
RC	STRAIN-Sprague-Dawley.				
RX	MEDLINE=20435277; PubMed=10978503;				
RA	Ritchie H.H., Wang L.-H.;				
RT	"The presence of multiple rat DSP-PP transcripts.";				
RL	Biochim. Biophys. Acta 1493:27-32(2000).				
RN	[2]				
RP	SEQUENCE OF 1-387 FROM N.A. AND SEQUENCE OF 18-26.				
RC	STRAIN-Sprague-Dawley.				
RX	MEDLINE=94148875; PubMed=8106414;				
RA	Ritchie H.H., Hou H., Vels A., Butler W.T.;				
RT	"Cloning and sequence determination of rat dentin sialoprotein, a novel dentin protein.";				
RL	J. Biol. Chem. 269:3698-3702(1994).				
RN	[3]				
RP	SEQUENCE OF 421-687 FROM N.A. (ISOFORM DPP-1).				
RC	STRAIN-Sprague-Dawley.				
RX	MEDLINE=96355511; PubMed=8702961;				
RA	Ritchie H.H., Wang L.-H.;				
RT	"Sequence determination of an extremely acidic rat dentin phosphoprotein.";				
RL	J. Biol. Chem. 271:21695-21698(1996).				
RN	[4]				
RP	SEQUENCE OF 1-19 FROM N.A.				
RC	STRAIN-Sprague-Dawley.				
RX	MEDLINE=99333695; PubMed=10403786;				
RA	Yamazaki H., Kunisada T., Miyamoto A., Tagaya H., Hayashi S.-I.;				
RT	"Tooth-specific expression conferred by the regulatory sequences of rat dentin sialoprotein gene in transgenic mice.";				
RL	Biochem. Biophys. Res. Commun. 260:433-440(1999).				
RN	[5]				
RP	SEQUENCE OF 29-33; 70-79; 93-109; 136-148; 162-188; 266-308; 398-423				
RC	AND 426-438, AND PHOSPHORYLATION OF SER-292 AND SER-298.				
RX	MEDLINE=21125612; PubMed=11042175;				
RA	Qin C., Cook R.G., Orlowski R.S., Butler W.T.;				
RT	"Identification and characterization of the carboxyl-terminal region of rat dentin sialoprotein.";				
RL	J. Biol. Chem. 276:904-909(2001).				
RN	[6]				
RP	TISSUE SPECIFICITY.				
RX	MEDLINE=98055479; PubMed=9395101;				

34	124.5	5.5	1411	1	TCOF_HUMAN	Q13428	homo sapien
35	124	5.4	675	1	MED1_MOUSE	P33215	mus musculus
36	124	5.4	706	1	SEM2_MACMU	O95196	macaca mula
37	124	5.4	1217	1	AF4_MOUSE	O88573	mus musculus
38	123	5.4	1395	1	SP41_YEAST	P38904	saccharomyc
39	123	5.4	2492	1	ATRX_HUMAN	P46100	homo sapien
40	122.5	5.4	1176	1	YOH8_YEAST	Q08236	saccharomyc
41	122	5.4	429	1	DR48_YEAST	P18899	saccharomyc
42	122	5.4	489	1	DMP1_RAT	P98193	rattus norv
43	121.5	5.3	598	1	CYLL_HUMAN	P35663	homo sapien
44	121.5	5.3	1165	1	YNF4_YEAST	P33950	saccharomyc
45	121.5	5.3	1960	1	TF20_HUMAN	Q9ug00	homo sapien

RA Ritchie H.H., Berry J.E., Somerman M.J., Hanks C.T., Bronckers A.L.,
 RA Horton D., Papagerakis P., Beraldi A., Butler W.T.,
 RT "Dentin sialoprotein (DSP) transcripts: developmentally-sustained
 RT expression in odontoblasts and transient expression in
 RT pre-ameoblasts";
 RL Eur. J. Oral Sci. 105:405-413(1997).
 CC -1- FUNCTION: DSP may be an important factor in dentinogenesis. DPP
 CC may bind high amount of calcium and facilitate initial
 CC mineralization of dentin matrix collagen as well as regulate the
 CC size and shape of the crystals.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms: DPP-1/PP240 (shown here), and
 CC DPP-2/PP171; may be produced by alternative splicing.
 CC -1- TISSUE SPECIFICITY: Specifically expressed in teeth, mainly in
 CC odontoblasts and transiently in pre-ameoblasts.
 CC -1- PTM: DSP is glycosylated.
 CC -1- CAUTION: Ref.2 sequence differs from that shown due to a
 CC frameshift in position 380.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AF247187; AAK6895.1; -
 DR EMBL: U02074; AAI18932.1; ALT_FRAME.
 DR EMBL: U63111; AAC52774.1; -
 DR EMBL: AF114987; AAD48588.1; ALT_TERM.
 KW Signal; Extracellular matrix; Glycoprotein; Sialic acid;
 KM Phosphorylation; Alternative splicing.
 FT SIGNAL 1 17
 FT CHAIN 18 687
 FT CHAIN 18 447
 FT CHAIN 448 687
 FT MOD_RES 57 57
 FT MOD_RES 226 226
 FT MOD_RES 253 253
 FT MOD_RES 278 278
 FT MOD_RES 292 292
 FT MOD_RES 298 298
 FT MOD_RES 315 315
 FT MOD_RES 319 319
 FT MOD_RES 329 329
 FT MOD_RES 337 337
 FT MOD_RES 345 345
 FT MOD_RES 366 366
 FT MOD_RES 55 55
 FT CARBOHYD 82 82
 FT CARBOHYD 128 128
 FT CARBOHYD 189 189
 FT CARBOHYD 312 312
 FT CARBOHYD 369 369
 FT VARSPLIC 567 635
 FT CONFLICT 74 74
 FT CONFLICT 564 564
 SEQUENCE 687 AA; 70179 MW; 9A845EED6AA31B63 CRC64;
 Query Match 6.98; Score 157; DB 1; Length 687;
 Best Local Similarity 21.58; Pred. No. 0.025;
 Matches 103; Conservative 67; Mismatches 167; Indels 142; Gaps 26;
 Oy 10 KENHNGIRMSITPKSTGNKGF-EDGDA-TSKLHDO--EYGAALLRNNOHIMGPVTA 65
 Db 127 ENSTRNGR-----GQVIAENAEAKSKYHGPHDPTKGLASDTSON--GDATL 176
 Oy 66 IKLGEENKENTPRNVNIIIPASNN-----YAKASKDKKKRPQDSQAQSPVSKS 117
 Db 177 V-----QENEPQ-----VAGSKNSTHVEVGTGSGVAQAQETTPQREGESSEN----- 218
 Oy 118 THRIQNDIYIKHLSKVKKIPDEFGSGYDLOERGNDISPFGSDGQPFMDINGKGEAT 177

Db 219 -----QGAEVTPSIGEGAGL-----DNTGSPSGNGIEDEEDTSGGCV 257
 Oy 178 GPDL-EGKDIOTGFAPSPSAEASHLDTKKPGVNEIPEREENGNTIGTDEAKEDAVD 236
 Db 258 GAGAGGRRSHGTE-----HEGSSGNN-----DNKQGSVTEDDSKED----- 301
 Oy 237 VSLVEGSSNDIMG--STNFKELPQRE--GNRYDAGSONAH--QGKVEFYPAPSPKRR 289
 Db 302 ----EESPNGRGDNTSSSEENGTEIEGDDGTQTTPDNQNSLPIEGGIIISQAEKPSQSGN 357
 Oy 290 EGSSDAEEST-NYNELPK-NGKSTKGYVDSNRNQTINEKORPPSKS---QCLP 344
 Db 358 QGLETEGSSNGKSSITKESGKLS--GSKDSNGHGMELDKRNSPKQESDKPQGAER 414
 Oy 345 S-----RGDNEIKNEMDSFNG-----PSHENITFH 371
 Db 415 SDTHNNMGHSRIGSSSDGHSYDFDDSDMOODPNSSDESNGSDGSDANSEKAIENG 474
 Oy 372 RKHYVPHRONSTRKNGKMPQCKGSGWGRPHSNRRFSRRRDDSSSSGSSSDGD 430
 Db 475 N-----HGDAVTSDESSDNGSDS--DSHAGEDDS--DDTSOTDSDSNGDDSE 521
 RESULT 2
 DSP_HUMAN STANDARD; PRT: 1253 AA.
 ID DSP_HUMAN
 AC Q9NZM4; Q95815;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Dentin sialoprophoprotein precursor [Contains: Dentin phosphoprotein
 DE (Dentin phosphophoryn) (DPP); Dentin sialoprotein (DSP)].
 GN DSP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP MEDLINE=20168992; PubMed=10706475;
 RA Gu K., Chang S.R., Ritchie H.H., Clarkson B.H., Rutherford R.B.;
 RT "Molecular cloning of a human dentin sialoprophoprotein gene";
 RL Eur. J. Oral Sci. 108:35-42(2000).
 RN [2]
 RP SEQUENCE OF 463-1253 FROM N.A.
 RC TISSUE=tooth;
 RX MEDLINE=99094526; PubMed=9879917;
 RA Gu K., Chang S.R., Slaven M.S., Clarkson B.H., Rutherford R.B.,
 RA Ritchie H.H.;
 RT "Human dentin phosphophoryn nucleotide and amino acid sequence";
 RL Eur. J. Oral Sci. 106:1043-1047(1998).
 RN [3]
 RP DISEASE.
 RX MEDLINE=21096971; PubMed=11175779;
 RA Zhang X., Zhao J., Li C., Gao S., Qiu C., Liu P., Wu G., Qiang B.,
 RA Lo W.H.Y., Shen Y.;
 RT "DSP mutation in dentinogenesis imperfecta Shields type II";
 RL Nat. Genet. 27:151-152(2001).
 RN [4]
 RP VARIANTS DENA39/DG11 THR-17 AND PHE-18.
 RX MEDLINE=21096982; PubMed=11175790;
 RA Xiao S., Yu C., Chou X., Yuan W., Wang Y., Bu L., Fu G., Qian M.,
 RA Yang J., Shi Y., Hu L., Han B., Wang Z., Huang W., Liu J., Chen Z.,
 RA Zhao G., Kong X.;
 RT "Dentinogenesis imperfecta I with or without progressive hearing loss
 RT is associated with distinct mutations in DSP";
 RL Nat. Genet. 27:201-204(2001).
 CC -1- FUNCTION: DSP may be an important factor in dentinogenesis. DPP
 CC may bind high amount of calcium and facilitate initial
 CC mineralization of dentin matrix collagen as well as regulate the
 CC size and shape of the crystals.
 CC -1- SUBCELLULAR LOCATION: Secreted.

SQ SEQUENCE 1253 AA: 126426 MW: DAE240553904ED4A CRR64;
 Query Match 6.7%; Score 152; DB 1; Length 1253;
 Best Local Similarity 23.5%; Pred. No. 0.1;
 Matches 123; Conservative 58; Mismatches 203; Indels 139; Gaps 27;

 QY 6 SISKENTHNLRLMSIVPKSTGNGKEFGDDAISLTHOEEFAGALLINMNHQINGPYTA 65
 Db ::::| | | | | | | | | | | | | | | | | | | | | | | | | | | |
 98 TLAAEGNEIEG-----WGGDTG-KAEYGHGCI--HKREKETA-----NGIGGVSTI 142
 QY 66 IKLLGEENKETPRNV-LNII PASNYAKAHSKD-----KKRPQ-----RDSOAQ 109
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 143 IDNAGATNRSTNTGNCTDKNTONGVDGA-GHNEDVAVVOEDGPQVAGSNNSNTDNDEITTE 201
 QY 110 KSPFKSSTHAIDINIDYLKHLKSVKKT-----PS-----DEGGGY 146
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 202 NSCHNEGTSSEITPOLINSKRNGTKEAEVTPGTGEDAGLDNSDGSPSGNADEDEDGSD 261
 QY 147 TDLEERGDNDISPPSGDQPFRKDIPGKQE---ATGPDLLEGKDIDTGAPGSEASTILD 202
 Db : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
 262 DEDEAGNGKDSNNKSNGQEGOD-HGKEDDHSSIGQNSDSKE---YYDEGCKDPHNE 316
 QY 203 -----TKKPQNELPERENGNGNTIGTRDELAK-----EADAVLSL- 239
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 317 VDGGRTSKSEENSAGIPEPDNGSQRI---EDQOKLNHRSKRVENKITKESETHAAGKSD 373
 QY 240 --VEGSNDIMGSTAF-KEL-PGREGNRYDAGSQNAHQEK-----VEFYHPAPSKE 286
 Db | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
 374 KGITIKPSSSNRRITKEVGKGNEG-KEDKQHGHILCKGVKVTQGEVYNIEGPQKSEP 432
 QY 287 KRREGSSDAEAESTN-----YNEIPLKNKGSGTRKGVDSNRRNQATLNEKORPPS----- 334
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 433 GNKGHGHTGSDSNSDGYSDYFDPKSMQGCDPPNSDESNCNDANDANSDDNNSSSRGDAS 492
 QY 335 ----KGRSGGLPIPSRGI-DNEIKENMDSEFGPSHENITIGRKXHYVPHQNSTRKNG 389
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 493 YNSESDKDNCGSDSKGAGEDDSDSTSOTNNSDSNGN-----GNNGMDNDK 539
 QY 390 MPQKGSWGROPHSNRRFRFSRRRDS-----SESSDGSSESSED 428
 Db ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 540 SDSGKGK-SDSDSDSDSDSNSSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSD 581

 RESULT 3
 DMPL_BOVIN STANDARD; -PRT; 510 AA.
 AC 095120;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Dentin matrix acidic phosphoprotein 1 precursor (Dentin matrix protein-1) (DMP-1).
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bos.
 NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=TOOTH;
 RX MEDLINE=97263953; PubMed=9109824;
 RA Hirst K.L., Ibaraki-O'Connor K., Young M.F., Dixon M.J.:
 RT "Cloning and expression analysis of the bovine dentin matrix acidic phosphoprotein gene."
 J. Dent. Res. 76:754-760(1997).
 CC -!- FUNCTION: COULD BE INVOLVED IN THE INDUCTION OF MINERALIZATION OF EXTRACELLULAR MATRIX AND IN DENTINOGENESIS
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN FETAL BRAIN, BONE AND TOOTH PARTICULARLY IN ODONTOBLAST, BUT NOT IN AMELOBLAST. NOT EXPRESSED IN LIVER AND SKIN.

FT MOD.RES 349 349 PHOSPHORYLATION (BY CK2) (POTENTIAL).
 FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 84 84 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 130 130 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 190 190 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 313 313 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 373 373 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 934 AA; 93901 MW; A618789D8A57249A CRC64;
 Query Match 6.5%; Score 147.5; DB 1; Length 934;
 Best Local Similarity 21.8%; Pred. No. 0.13;
 Matches 105; Conservative 63; Mismatches 163; Indels 151; Gaps 27;
 QY 10 KENTHGLRMSIYPKSTGNKF-EDGDDAISKLHDQ-----EEYGAALLIRNNQHIIMPV 63
 DB 129 ENSTANGIRSQV-----GIVENAEAESESVHGQAGMTKSGADVSQN-----GDA 175
 QY 64 TAIKLLGEENKENTPRNVLIIPASMYAAH-----SKDKKQPRD---SQAQSPYKSK 116
 DB 176 TLV-----QENPEPPA-SIKNSTNHEAGIHSGVATHTTTPQREGLGSENOGTEV--- 224
 QY 117 STHRIOHNDYLLKHLKVKKIPSPDFEGSGYTDQERNDNISPFSGDGPFKDIPKGEA 176
 DB 225 -----TPSIGEDAGL-----DDTGSP-SGNGVEDEDEDTGSDG 257
 QY 177 TGPDL-EGKDIQTGFAPR-----SEAESTHLDTKRP-----GYNEIDPEREE 216
 DB 258 EGAEAGCGRESHDGTGKGOGSGHGMTHRGOGSVSTEDDDSKQEGEFGPHNGDENSEE 317
 QY 217 N-----GGNTIGTRD-----ETAKKADAVUSLVE-----GSNDIMG-----STNEKELP 256
 DB 318 NGVEEGDSTQATQDKKEKLSKPRDRAEGGILISQSEACPSKSDDOGIEGPKNGKNSII 377
 QY 257 GREGNRYDAGSQ--NAHQGVKEFHPAPSKERKEGSSPAESTVNEIIPKNGKSTRK 314
 DB 378 TKSSGL-SSSKDSNGHGQ-VELDKRNP-----KQGESKPKQ-----GTAEK 418
 QY 315 GVDHSRNQATLNEKORFPKSGKSGQ--LPIPSRLGDIENKEMDSFNGPSHENITTHG 371
 DB 419 SAHNSNLGHS-----RIGSSNSDGHDSYEFDEDSQMGDDPKSPESNSDSESD----- 467
 QY 372 RKHYHYPHNONSTRNKMGQKSGMSGRPHS---NRFRSRRRDDSSSDSGSSSESD 428
 DB 468 -----TNSSEANESGSGDASYTSPDESSDDDDSDSHAGEDSDSDSGDGDSDSN 518
 QY 429 GD 430
 DB 519 GD 520
 RESULT 5
 CYL1_BOVIN STANDARD; PRT; 667 AA.
 ID CYL1_BOVIN P35662;
 AC P35662;
 DT 01-JUN-1994 (rel. 29, Created)
 DT 01-JUN-1994 (rel. 29, Last sequence update)
 DT 01-NOV-1997 (rel. 35, Last annotation update)
 DE Cyclicin I (Multiple-band polypeptide I).
 GN CYL1 OR CYL.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OC NCB1_TaxID=9913;
 OX NCB1_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE-Testis.
 RX MEDLINE=93359502; PubMed=8354692;
 RA Hess H., Heid H., Franke W.W.;
 RT "Molecular characterization of mammalian cyclin, a basic protein of
 the sperm head cytoskeleton."
 RT J. Cell Biol. 122:1043-1052(1993).
 CC -I- FUNCTION: POSSIBLE ARCHITECTURAL ROLE DURING SPERMATOGENESIS. MAY

CC BE INVOLVED IN SPERMATID DIFFERENTIATION.
 CC -I- SUBCELLULAR LOCATION: CALYX; SPERM HEAD CYTOSKELETAL STRUCTURE.
 CC -I- TISSUE SPECIFICITY: TESTIS.
 CC -I- DEVELOPMENTAL STAGE: SPECIFIC TO LATE SPERMATOGENESIS.
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: 222779; CAA80456.1; -
 DR PIR: S35913; S35913.
 DR PIR: A40713; A40713.
 KM Cytoskeleton; Structural protein; Repeat; Sperm. Spermatogenesis.
 FT DOMAIN 287 569 9 APPROXIMATE TANDEM REPEATS.
 FT REPEAT 287 305 1.
 FT REPEAT 306 337 2.
 FT REPEAT 338 368 3.
 FT REPEAT 369 405 4.
 FT REPEAT 406 442 5.
 FT REPEAT 443 475 6.
 FT REPEAT 476 517 7.
 FT REPEAT 517 547 8.
 FT REPEAT 548 569 9.
 FT DOMAIN 617 667 PRO-RICH.
 SQ SEQUENCE 667 AA; 74817 MW; CBF66E462243D91 CRC64;
 Query Match 6.4%; Score 146; DB 1; Length 667;
 Best Local Similarity 20.5%; Pred. No. 0.11;
 Matches 97; Conservative 70; Mismatches 190; Indels 116; Gaps 22;
 QY 22 YPKST---GNKGPEDDDAIS-KLHQEYGAALLIRNNQHIIMPVTAIKLLGEENKENT 77
 DB 55 FPKSQPGGNKRLRPEIOTVYPRHD-----KRNLDELQKPAHIWIRHSLRKKFOS 105
 QY 78 PRNVLIIT---PASMNY-----AKANSK---DKK-----KPK 103
 DB 106 PS--IMLIYRQASFRHPYTHITHSKKAKSKYKDKKLTALKIKKIDTGPHEVDEKPR 163
 QY 104 ROSQAQSPYK-----KSTRIOHND-----YLKLSKVKKIPSPDESGGYTD 148
 DB 164 RRRKADKTPKSSKSHGSQLSKSKSKSETNPESDJSVSIKHKKRKRKSD---SKEND 220
 QY 149 LQERGNDISPFSGDQPRKIDPGKEATGPDEGDIQTGF---AGPSEASTHLDTKK 205
 DB 221 FESTSTKYS-----KSKNNSDAVSETCSKNSNVGLVHLGESDAESMEFDMVL 271
 QY 206 PGYNEIPEERENGNTIGTRDETAKKADAVDSLVGSDNDMGST-NFKELPEREGNRVD 264
 DB 272 KNTSQNSKKPKTKD--AKKDAKGGSDAESVDSKDAKKDKATKDTIKGAKKIDESTD 329
 QY 265 A--GSQNAHQGVKEFHPAPSKERKEGSSDAESTVNEIIPKNGKSTRKGVDSNR 321
 DB 330 AESGDSKDAKKGKE-----SKDKKDAKRAADAESGD-SKAAKDSKGGKDKSK 382
 QY 322 NQATLNEKORFP-----KKSQGLPIPSRLGDIENKEMDSFNGPSHENITTHGRYHYV 377
 DB 383 DNKKKAKKDAESTDAESDSDSKAKKSK-----GKKDSDSKDKDKK-----AKKDAES 432
 QY 378 PHRONNSTNRNKGMPGKSGMSGRPHSNRFRSRRRDDSSSDSGSSSESDG 430
 DB 433 TDAESGDSKNAKKDKSKG-----KKDKKKKAKKADAVSTDAESSEGD 475
 RESULT 6
 DMP1_HUMAN STANDARD; PRT; 513 AA.
 ID DMP1_HUMAN Q13316; Q43265;
 AC Q13316; Q43265;
 DT 30-MAY-2000 (rel. 39, Created)
 DT 30-MAY-2000 (rel. 39, Last sequence update)
 CC

```

DT 16-OCT-2001 (Rel. 40, last annotation update)
DE Dentin matrix acidic phosphoprotein 1 precursor (Dentin matrix
DE protein-1) (DMP-1).
GN DMP1.
OC Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE: Molari;
RX MEDLINE=97321043; PubMed=9177774;
RA Hirst K.L., Simmons D., Feng J., Apelin H., Dixon M.J., McDougall M.;
RT "Elucidation of the sequence and the genomic organization of the human
RT dentin matrix acidic phosphoprotein 1 (DMP1) gene: exclusion of the
RT locus from a causative role in the pathogenesis of dentinogenesis
RT Imperfecta type II."
RT Genomics 42:38-45(1997).
RL [2]
RN SEQUENCE FROM N.A. (ISOFORM 2).
RA McDougall M., Juan X., Simmons D., Feng J.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 462-513 FROM N.A., AND GENE MAPPING.
RX MEDLINE=96163890; PubMed=8586437;
RA Apelin H.M., Hirst K.L., Crosby A.H., Dixon M.J.;
RT "Mapping of the human dentin matrix acidic phosphoprotein gene (DMP1)
RT to the dentinogenesis imperfecta type II critical region at chromosome
RT 4q21."
RL Genomics 30:347-349(1995).
CC -1- FUNCTION: COULD BE INVOLVED IN THE INDUCTION OF MINERALIZATION OF
CC EXTRACELLULAR MATRIX.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: 1. (SHOWN HERE) AND 2; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN TOOTH PARTICULARLY IN
CC OONTOBLAST, AMELOBLAST AND CEMENTOBLAST.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U89012; AAC51332.1; -
DR EMBL: U34037; AAA97602.1; -
DR EMBL: U65378; AAB87728.1; -
DR Genew: HGNC:2932; DMP1.
DR MIM: 600980; -
KW Extracellular matrix; Signal; Alternative splicing.
FT SIGNAL 1 16
FT CHAIN 17 513
FT CARBOHD 25 25 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHD 324 324 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHD 345 345 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHD 351 351 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHD 413 413 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHD 426 426 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHD 467 467 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT SITE 364 366 CELL ATTACHMENT SITE (POTENTIAL).
FT VASPLIC 46 61 MISSING (IN ISOFORM 2).
FT CONFLICT 69 69 S->C (IN REF. 2).
SO SEQUENCE 513 AA; 55782 MW; 2C1FDE319A5D106F CMC64;
Query Match 6.2%; Score 141.5; DB 1; Length 513;
Best Local Similarity 21.1%; Pred. No. 0.14;
Matches 94; Conservative 76; Mismatches 181; Indels 95; Gaps 20;

```

```

QY 82 INIPASMNVAKASHKDKKPPQPSDAQKSPYKSKSTHRIQHNIDYIKLHLSKVKKIPSD 141
DB 141 DDTIOAS-----ESAPQGDSDADTTSFS-----RELINEDRVSKP 178
QY 142 EGGGYTDLOER-----GDNDISPPSGDQPPKIPGKGEATPGDLCKDIQTGFACGSE 195
DB 179 EGGDSTQSESEEHVWGCGSDGESHGDSLELD-----EGMQSDPEISIRSRGMSR 231
QY 196 AESTHLDTKKPGYN-EIPIRENGN-----TIGRDETAKEADAVDV-----SLVEG 242
DB 232 MNSGAMKSKSEGENSEQANTDQSGSOLLEHPKRTFKKSISEDDRSLELDNNTMEV 291
QY 243 SNDIGSTNFKEL-----PGRE--GNRYDAGSONAHQGVKEPHYPAPKEREKKEG--SSD 294
DB 292 KSDSTENSNSDTGLSGQRDRSKDQSDSKENLSQ--EESQNVDPGSESSQEANLSQ 349
QY 295 AESTNVEIPIRNGKS-----TRKGVDSNNQATLNKQKFPKSGKQGLPIFSRGIDN 350
DB 350 ENSSESGEEVYSESGKDPDPTTSYVEDQEDSDSESDSHTLSKSES--REQADS 406
QY 351 EIKNEMSEFNGPSENITTHGRKRYHYVPHRONSTR-----NKGMFGKGSWCRQPHS 403
DB 407 E-SSBSLNFSESPDS-----PEDENSSQEGLOSHSSASQSESHSEEDDS 454
QY 404 NRRFSRRRDSSESSDSSSESDG 429
DB 455 DSQSSRSKEDSN-STESKSSSEEDG 479
RESULT 7
MLH_TETTH STANDARD; PRT; 633 AA.
AC P40631;
ID 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, last annotation update)
DE Micronuclear linker histone polypeptide (MTC LH) [Contains:
DE Micronuclear linker histone-alpha; Micronuclear linker histone-beta;
DE Micronuclear linker histone-delta; Micronuclear linker histone-gamma].
GN MLH.
OS Tetrahymena thermophila.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC Tetrahymenina; Tetrahymena.
OX NCBI_TaxID=5911;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=CU401;
RX MEDLINE=94088506; PubMed=8264578;
RA Wu M., Allis C.D., Sweet M.T., Cook R.G., Thatcher T.H.,
RA Gorovsky M.A.;
RT "Four distinct and unusual linker proteins in a mitotically dividing
RT nucleus are derived from a 71-kilodalton polypeptide, lack p34cdc2
RT sites, and contain protein kinase A sites."
RT Mol. Cell. Biol. 14:10-20(1994).
CC -1- SUBCELLULAR LOCATION: NUCLEAR; MICRONUCLEI.
CC -1- PTM: ALL FOUR HISTONES ARE PROCESSED FROM THE PRECURSOR MOLECULE.
CC THEY ARE PHOSPHORYLATED IN GROWING AND DIVIDING CELLS BUT NOT IN
CC NONGROWING (STARVED) CELLS. THE N-TERMINAL OF ALPHA AND DELTA IS
CC BLOCKED.
CC -1- SIMILARITY: CONTAINS 2 HMG BOXES.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M87306; AAC18874.1; -
DR InterPro: IPR000910; HMG_12_box.
DR Pfam: PF00505; HMG_box; 1.

```

```

RT   region segments."
RL   Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
RN   [4]
RX   ALTERNATIVE SPLICING.
RA   MEDLINE=94061824; PubMed=8242607; Oshima M., Sugiyama H., Kitagawa K., Takeeto M.;
RT   "APC gene messenger RNA: novel isoforms that lack exon 7.";
RL   Cancer Res. 53:5589-5591(1993).
CC   - FUNCTION: TUMOR SUPPRESSOR. ALLOWS THE RAPID TURNOVER OF BETA-CATENIN. APC ACTIVITY CORRELATED WITH ITS PHOSPHORYLATION STATE
CC   - ILLUSTRATES THE DOWN-REGULATION OF CYTOPLASMIC BETA-CATENIN (BY SIMILARITY).
CC   - SUBUNIT: FORMS HOMODIGOMERS AND ASSOCIATES WITH CATEININS (BY SIMILARITY).
CC   - ALTERNATIVE PRODUCTS: 4 ISOFORMS; 1 (SHOWN HERE), 2, 3 AND 4; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   - TISSUE SPECIFICITY: EXPRESSED IN LIVER, SPLEEN, KIDNEY, HEART, LUNG, BRAIN, STOMACH, INTERSTING, TESTIS AND OVARY.
CC   - PTM: PHOSPHORYLATED BY GSK-3B (BY SIMILARITY).
CC   - SIMILARITY: CONTAINS 7 ARM REPEATS.
-----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation CC   the European Bioinformatics institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sdb.ch/announce or send an email to license@isb-sdb.ch).
-----
DR   EMBL; M88127; AAB59632.1; -
DR   EMBL; U02937; AAA03443.1; -
DR   HSSP; 002248; 3BCF.
DR   MGD; MG1:88039; APC.
DR   InterPro: IPRO00225; Armadillo.
DR   Pfam: PF00514; Armadillo-seg; 6.
DR   SMART; SM00185; ARM; 5.
DR   PROSITE; P550176; ARM_REPEAT; 1.
KW   Anti-oncogene; phosphorylation; Alternative splicing; Repeat; Coiled coil.
FT   DOMAIN          1      61      COILED COIL (POTENTIAL).
FT   FT              125     245     COILED COIL (POTENTIAL).
FT   DOMAIN          1      728     LEU-RICH.
FT   REPEAT          451     493     ARM 1.
FT   REPEAT          503     545     ARM 2.
FT   REPEAT          546     589     ARM 3.
FT   REPEAT          590     636     ARM 4.
FT   REPEAT          637     681     ARM 5.
FT   REPEAT          682     723     ARM 6.
FT   REPEAT          724     765     SER-RICH.
FT   DOMAIN          739     2634    ASP/GLU-RICH (ACIDIC).
FT   DOMAIN          1130    1156    ASP/GLU-RICH (ACIDIC).
FT   DOMAIN          1556    1575    ASP/GLU-RICH (ACIDIC).
FT   DOMAIN          1864    1891    HIGHLY CHARGED.
FT   VARSPLIC        243     276     MISSING (IN ISOFORM 2 AND ISOFORM 4).
FT   VARSPLIC        310     410     MISSING (IN ISOFORM 3 AND ISOFORM 4).
FT   VARIANT          120     120     T -> A (IN STRAIN CAST/EI).
FT   VARIANT          493     493     V -> I (IN STRAIN CAST/EI).
FT   VARIANT          797     797     Y -> F (IN STRAIN CAST/EI).
FT   VARIANT          1330    1330    A -> T (IN STRAIN CAST/EI).
FT   VARIANT          1618    1618    A -> S (IN STRAIN CAST/EI).
FT   VARIANT          2294    2294    G -> A (IN STRAIN CAST/EI).
FT   VARIANT          2496    2496    H -> Q (IN STRAIN CAST/EI).
FT   VARIANT          2523    2523    T -> A (IN STRAIN CAST/EI).
FT   VARIANT          2813    2813    T -> S (IN STRAIN CAST/EI).
SQ   SEQUENCE      2845 AA; 311086 MW; 145CA73CF570A499 CRC64;
Query Match      6.2%; Score 140.5; DB 1; Length 2845;
Best Local Similarity 21.3%; Pred. No.1.3;
Matches 108; Conservative 67; Mismatches 212; Indels 121; Gaps 23;

```

QY 69 LGSEKENTPRNNVILIPASNNYAKHAKSKKKKPPORDS--QAAQSPYKSKSTRICHNT 125
 Db 2152 PDSEKPTPTS---NKGRILKPGSEKSTLEAKKIESEKKGKIGKVKYKSLITKIRSNS 2207
 QY 126 DYKHLKSK--VKKIPSEDESGYDIO--EKGDNDISFGSDGDPFKDIPKGEATGPD 181
 Db 2208 EISSOMKOPPLPTNMPSTISRGHTMHIPELRNSSSTSVSKKGPPLKTPAKSKSESGGA 2267
 QY 182 EKGDIQTGFGAPSE-AESTHLDTKKPGYNEIPEEEENGNTIGRDETAKEADAVD--- 237
 Db 2268 TTSPTGCTPAKSKSELSPITRQTSQSGSNK-----GSSRSGSDSPSPRPTQOPLSRP 2320
 QY 238 -----SLVSGSDNDIMSTNFKELPGREGNRVAGSOMNAQKVEFYHPAPSKERKE 290
 Db 2321 MOSPGRNMSISPGRNISPPNKLSQLP-RTSSPSTASTSKSGSGKMSY---TSPGQLSQ 2376
 QY 291 G-SSDAESTYVNEIPK-----NGKSTRK-----GVDSHNR-- 321
 Db 2377 NITKASLSKNAASIPRESASKGLNOMNGSKKKKVELRMSSTKSGSESDSERPA 2436
 QY 322 --NQAT-----LNEKQRPSPKSKSGGLPIPSRGLDNE--IKNEMDSFNGPSH 364
 Db 2437 LVROSTFIKEAPSPILRRKLEESASFESLSPSRDPSPTRSQAQTPVLSPLPDMSLSTH 2496
 QY 365 ENITTHG-RKYHYVPHRQNNSTRNKGMPQK-----GSMGRO--PH 402
 Db 2497 PSVQAGMKR--LPPNLSPITTYNDGRPTKRHDIRKSHSESPRLPINRAGTKWKEHSHKH 2554
 QY 403 SNR--RFSRRRDSSESDSGSSSED 428
 Db 2555 SSSLPRVSTWRRTGSSSILSSSESSE 2582

RESULT 9
 NKX1_RAT STANDARD: PRF: 1181 AA.

AC 090ZM6: 062932: 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Sodium/potassium/calcium exchanger 1 precursor (Na(+)/K(+)/Ca(2+)-
 exchange protein 1) (Retinal rod Na-Ca+K exchanger).
 GN SLC24A1 OR NCKX1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBL_TaxID=101116;
 RN [1]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RC STRAIN=Sprague-Dawley; TISSUE=Eye;
 RX MEDLINE=2021735; PubMed=10751314;
 RA Poon S., Leach S., Li X.-F., Tucker J.E., Schneckamp P.P.M.,
 RA Lytton J.,
 RT "Alternatively spliced isoforms of the rat eye
 sodium/calcium/potassium exchanger NCKX1.";
 RL Am. J. Physiol. 278:C651-C660(2000).
 RN [2]
 RP SEQUENCE OF 1067-1155 FROM N.A.
 RA White K.E., Gesek F.A., Friedman P.A.;
 RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Critical component of the visual transduction cascade,
 controlling the calcium concentration of outer segments during
 light and darkness. Light causes a rapid lowering of cytosolic
 free calcium in the outer segment of both retinal rod and cone
 photoreceptors and the light-induced lowering of calcium is caused
 by extrusion via this protein which plays a key role in the
 process of light adaptation. Transports one Ca(2+) and one K(+) in
 exchange for four Na(+).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- ALTERNATIVE PRODUCTS: 4 isoforms; 1 (shown here) 2, 3 and 4; are
 produced by alternative splicing.
 CC -1- TISSUE SPECIFICITY: Highly expressed in the eye.

CC -1- SIMILARITY: BELONGS TO THE SLC24A FAMILY OF TRANSPORTERS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AF176688; AAD53121.1; -
 CC EMBL: U49235; AAB37753.1; -
 CC InterPro: IPR004817; K_NaCaexchng.
 CC InterPro: IPR004481; K_NaCaexchng.
 CC InterPro: IPR004837; NaCa_Exmemb.
 CC Pfam: PF01699; NaCa_Ex_3.
 CC TIGRFAMs: TIGR00367; K_NaCaexchng-rel; 1.
 CC TIGRFAMs: TIGR00927; 2A1904; 1.
 CC Vision; Transport; Antiport; Symport; Calcium transport;
 CC Transmembrane; Glycoprotein; Phosphorylation; Signal; Repeat;
 CC Alternative splicing.
 KW SIGNAL
 FT 1 38
 FT CHAIN 39 1181
 FT DOMAIN 39 419
 FT TRANSMEM 420 440
 FT DOMAIN 441 464
 FT TRANSMEM 465 485
 FT DOMAIN 486 491
 FT TRANSMEM 492 512
 FT DOMAIN 513 519
 FT TRANSMEM 520 544
 FT DOMAIN 545 552
 FT TRANSMEM 553 569
 FT DOMAIN 570 989
 FT TRANSMEM 990 1010
 FT DOMAIN 1011 1017
 FT TRANSMEM 1018 1038
 FT DOMAIN 1039 1053
 FT TRANSMEM 1054 1074
 FT DOMAIN 1075 1092
 FT TRANSMEM 1093 1113
 FT DOMAIN 1114 1121
 FT TRANSMEM 1122 1142
 FT DOMAIN 1143 1150
 FT TRANSMEM 1151 1171
 FT DOMAIN 1172 1181
 FT REPEAT 461 501
 FT REPEAT 1061 1092
 FT DOMAIN 730 905
 FT REPEAT 730 741
 FT REPEAT 742 754
 FT REPEAT 755 766
 FT REPEAT 767 778
 FT REPEAT 779 791
 FT REPEAT 792 804
 FT REPEAT 805 817
 FT REPEAT 818 830
 FT REPEAT 831 843
 FT REPEAT 844 856
 FT REPEAT 857 869
 FT REPEAT 870 881
 FT REPEAT 882 893
 FT REPEAT 894 905
 FT MOD_RES 952 974
 FT MOD_RES 625 625
 FT CARBOHYD 271 271
 FT VARSPLIC 598 710
 FT VARSPLIC 616 710
 FT VARSPLIC 652 679
 SQ SEQUENCE 1181 AA; 129980 MW; B063C1C1193696AE CRC64;
 Query Match 6.1%; Score 138; DB 1; Length 1181;
 Best Local Similarity 22.1%; Pred. No. 0.62;


```

Matches 89; Conservative 58; Mismatches 188; Indels 68; Gaps 21;

QY 39 SKLHDEEYGAALIRNMOMHMGVTAIKLIGEEKNENTRNVTNITPASNVAKAH-SK 97
Db 627 ASLHN-----SIRSTTHLM--LHSLDPLGEARPSKDKOESLNQPARVLPQTKASSS 678
QY 98 DKKKPORDSQAOKSPVKS-----KSTHRIQNHIDYKLHLSKKKIPSPFEGSGYDQ 150
Db 679 DEEPFALPLAVTVPAPAPEDKQDQEDPGQEDVDIAEHRGDMTGEGEREETAEKKD 738
QY 151 ERQDNDSPPSGDQPKP-DIPCKGEATG-PDLEKQIQIGFAPGSAESTHLDTKKPGY 208
Db 739 EEGETE-AERKEDQEEETETKGEKEGESEKDEQ---EGETAEKEDHE--GE 792
QY 209 NEIPERE-ENGNT--IGTRDETAKEDAVDSL-VEGSDINGSTFKELPREGNRVD 264
Db 793 TEAEKVEHEGETEAGCTEDDEDEGETEAGKEVEDEGETEABD---KEVEHEVETAE 848
QY 265 AGSOMAHQKVEFHYPPAPSKERKEGSSDAASTNY-NEIPKNGKSTKGYDHSNRQ 323
Db 849 RKEFN-HEGETE-----AEKADHEGETEAGVEHQGTAEAGK-----VEHEGETE 896
QY 324 ATLNEGRFSPKSGSGLPLPSGLDNEIKNEMDSFNGPHEHNTIHKRYHYVPHRNN 383
Db 897 A--GEKDEHGGSETQ-----ADTEYKDG-----EGEAEN-----AEDQCT 933
QY 384 STANKGMPGKSGWGRPHSNRRFSRRDSSSSSDSSSE 426
Db 934 AGEKAGDGGSGDGDSEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE 976

RESULT 10
SERI_BOMMO STANDARD; PRT; 389 AA.
AC P07856;
DT 01-AUG-1988 (Rel. 08, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Sericin precursor (Silk gum protein).
OS Bombyx mori (Silk moth).
OC Insecta; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Dityrsta; Bombycoidea; Bombycidae; Bombyx.
OX NCBI_TaxID=7091;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Middle silk gland;
RX MEDLINE=83082839; PubMed=6294094;
RA Okamoto H., Ishikawa E., Suzuki Y.,
RT "Structural analysis of sericin genes. Homologues with fibroin gene
RT in the 5' flanking nucleotide sequences."
RT J. Biol. Chem. 257:15192-15199(1982).
RN [2]
RP SEQUENCE OF 317-354 FROM N.A.
RX MEDLINE=87076763; PubMed=3024742;
RA Michaille J.J., Couble P., Prudhomme J.-C., Garel A.;
RT "A single gene produces multiple sericin messenger RNAs in the silk
RT gland of Bombyx mori."
RC Biochimie 68:1165-1173(1986).
RT -1- FUNCTION: PROVIDES THE SILK FIBROIN THREAD WITH A STICKY COATING.
CC ACTS AS A CEMENT BY STICKING SILK THREADS TOGETHER.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: PRODUCED EXCLUSIVELY IN THE MIDDLE (MSG)
CC SECTION OF SILK GLANDS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL: J01029; -; NOT_ANNOTATED_CDS.
DR EMBL: J01030; -; NOT_ANNOTATED_CDS.
DR EMBL: J01031; -; NOT_ANNOTATED_CDS.
DR EMBL: J01032; -; NOT_ANNOTATED_CDS.
DR EMBL: J01033; -; NOT_ANNOTATED_CDS.
DR EMBL: J01034; -; NOT_ANNOTATED_CDS.
DR EMBL: M26102; AAA27843.1; -;
KW SILK; Signal.
FT SIGNAL 1 15
FT CHAIN 16 389
SQ SEQUENCE 389 AA; 38820 MW; BA605BC0305EAF19 CRC64;

Query Match 6.0%; Score 136; DB 1; Length 389;
Best Local Similarity 21.1%; Pred. No. 0.22;
Matches 71; Conservative 55; Mismatches 166; Indels 44; Gaps 11;

QY 115 SKSTHRIQNHIDYKLHLSKKKIPSPFEGSGYDQDERGQNDISPPSGDQPKDIPKG 174
Db 41 SESSYLNKQNDISAGAHRAKSVESQDQSKYTS---GPEGVST-SGNSQNTKDSKQAI 95
QY 175 EATGPLEGKDIQTGFAGPSEASTHLDTKKPGYNEIPEREENGNTIGTRDETAEADA 234
Db 96 ISGCTKSSNSNVQSDKKSASQSSSSSRSSQESAYS-----SSSSSTESSSSSSRAAS 150
QY 235 VDVSLVEGNS-DIMGSTNFKELPREGNRVDAGSQAHAQKVEFHYPPAPSKERKEGSS 293
Db 151 TDASSNTDSNSNAGSST-----SGGRRTYGYSSNSRDGVS---STGSSSNTDSNS 200
QY 294 DAAESTN-----YNEIPKNGKSTKGYDHSNRQATLNEGRFSPKSGSGLPIPSR 346
Db 201 MAGSSTSGSSSTGYGSSNSRDGVSFTGSSSNTDSNSNVGSR-----SGSSSHDSSK 256
QY 347 GLDNEIKNEMDSFNGPHEHNTI--THG--KRYHYVPHRNNSTRNKGMPOK----- 394
Db 257 SRDENVTGSSSNTDSNSNVGSSSTSGGRRTYGYSSNSRDGVSSTGSSSNTDSNSNV 316
QY 395 GSWGRPHSNRRFSRRDSSSSSDSSSGSSSDG 430
Db 317 GSSTSGSSSTGYGSSNSRDGVSSTGSSSNTDSNSN 352

RESULT 11
AF4_HUMAN STANDARD; PRT; 1210 AA.
AC P51825;
DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE AF-4 protein (Proto-oncogene AF4) (FEL protein).
GN MLF2 OR AF4 OR FEL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93281633; PubMed=8506309;
RA Nakamura T., Alder H., Gu Y., Prasad R., Canaan O., Kanada N.,
RA Gale R.P., Lange B., Crist W.M., Nowell P.C., Croce C.M.,
RA Canaan E.;
RT "Genes on chromosomes 4, 9, and 19 involved in 11q23 abnormalities in
RT acute leukemia share sequence homology and/or common motifs."
RT Proc. Natl. Acad. Sci. U.S.A. 90:4631-4635(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93184301; PubMed=8443374;
RA Morrissey J., Tkachuk D.C., Milatovich A., Francke U., Link M.,
RA Cleary M.L.;
RT "A serine/proline-rich protein is fused to HRX in t(4;11) acute
RT leukemia."
RT Blood 81:1124-1131(1993).
RN [3]
RP -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -----

```


RESULT 13		
REPT_MOUSE	STANDARD:	PRT: 1130 AA
ID REPT_MOUSE		
AC P97317;		
DT 01-NOV-1997	(Rel. 35, Created)	
DT 01-NOV-1997	(Rel. 35, Last sequence update)	

```

DE DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Reptin.
GN RPTN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MRI; TISSUE=Skin;
RX MEDLINE=97422611; PubMed=9268637;
RA Krieb P., Schuppler M., Koesters R., Mincheva A., Lichter P.,
RA Mairs F.;
RT "Reptin (Rptn), a new member of the 'fused gene' subgroup within the
RT S100 gene family encoding a murine epidermal differentiation
RT protein."
RL Genomics 43:339-348(1997).
CC -!- FUNCTION: NOVEL POTENTIAL PRECURSOR PROTEIN OF THE CORNIFIED CELL
CC ENVELOPE.
CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE S-100
CC FAMILY.
CC -!- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X99251; CAA67624.1; -.
DR HSSP; P02633; 1IG5.
DR MGD; MGI:1099055; Rptn.
DR InterPro; IPR001751; CABP-S100.
DR InterPro; IPR02048; EF-hand.
DR Pfam; PF00036; ethand; 1.
DR Pfam; PF01023; S-100; 1.
DR ProDom; PD003407; CABP-S100; 1.
DR PROSITE; PS00018; EF_HAND; 1.
DR PROSITE; PS00303; S100_CABP; 1.
DR Repeat; Calcium-binding.
DR KW DOMAIN 1 91 S-100 LIKE.
FT CA_BIND 22 33 EF-HAND 1 (LOW AFFINITY) (POTENTIAL).
FT FT 62 73 EF-HAND 2 (HIGH AFFINITY) (POTENTIAL).
FT CA_BIND
FT SEQUENCE 1130 AA; 129884 MW; 7D97BDD25151918D CRC64;
Query Match 5.9%; Score 135.5; DB 1; Length 1130;
Best Local Similarity 19.1%; Pred. No. 0.82; Mismatches 190; Indels 117; Gaps 19;
Matches 87; Conservative 61;
QY 8 SNKENTHNGLRMSIYPKSTGKNGKGFEGDDAISKLHDEEYGAALIRNMCHIMGPVAK 67
QY 100 SQKEHDEGTGSHHFSESTGRQ-----HROYEE-----RRNSHH----- 135
QY 68 LLGEKNKENTPRNVLIITIPASMYAKAKSKKKRQNDQAOKSPV--KSKSTHRIQHN 124
QY 136 -----NDSGGLQANQH-----DOSQRDKKSEKHDDHPDHCQSETRFGDSHYGSEKD 185
QY 125 IDYLKHLSKYKRIPISEFGS-----GYTDLQE--RGDNDISPSGDCQPKDIR-GKGE 175
QY 186 TDVSSDQSE-----SDNESSSSQRLCYKSSHEQPKGGYVFFALSQSKNPQAHYQSK 240
QY 176 ATGDLGKGDIDQNGFAPRSEAEETHLDTKKPKGYNEIIEREENGNTIGTDETAKEADAV 235
QY 241 TSGG--QSHGQSGRFKKDYSSTQSOEEDSYEYGSQHOKSNSQSTER----- 288
QY 236 DVSLEVGSDIMGISTNFKELPGREGNVADGSONAH-----QKVEEYHPAPSEK--- 287
QY 289 -----QGNQSYQGTNKK-----GHSSYHQGTGQGG--SHYQGKGRKDSFQ 330
QY 288 -----REGSSDAAESTNYEIKNGKSGSTRKGYDVSNNRQATLNEKQRPSPKSGKSGGLP 342

```

RN SEQUENCE FROM N.A.
 RP STRAIN-Swiss Webster; TISSUE:molar;
 RC MEDLINE=96184257; PubMed=9525343;
 RX McDougall M., Gu T.T., Luan X., Simmons D., Chen J.,
 RA "Identification of a novel isoform of mouse
 RT spatial expression in mineralized tissues".
 RL J. Bone Miner. Res. 13:422-431(1998).

Accession	Gene	Protein
CC	EMBL: U65020; AAB93764.1; -	
DR	EMBL: AJ242625; CAB59629.1; -	
DR	MCD; MGI:94910; Dmpl1	
KW	Extracellular matrix; Signal.	
FT	SIGNAL	POTENTIAL.
FT	CHAIN	DENTIN MATRIX ACIDIC PHOSPHOPROTEIN 1.
FT	DOMAIN	POLY-PRO.

Query Match	Score	DB	Length
Best Local Similarity	19.68;	Pred. No. 0.36;	503;

RESULT 15 :		
PVDR_PLAYS	STANDARD;	PRT; 1070 AA.
ID	P22850	
AC	P22850	
DA	01-MAY-1991	(Rel 19, Created)
DA	01-MAY-1996	(Rel 33, Last sequence update)
DA	30-MAY-2000	(Rel 39, Last annotation update)

DE Duffy receptor precursor (Erythrocyte binding protein).
GN PVDR.
OS Plasmodium vivax (strain Salvador I).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID-126793;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-91187056; PubMed-1849231;
RA Fang X., Kaslow D.C., Adams J.H., Miller L.H.;
RT "Cloning of the Plasmodium vivax Duffy receptor.";
RL Mol. Biochem. Parasitol. 44:125-132(1991).
CC -1- FUNCTION: BINDS TO THE HUMAN ERYTHROCYTES DUFFY BLOOD GROUP
CC DETERMINANT.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: HIGH, TO P.KNOWLEDST DUFFY RECEPTORS.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sb.ch/announce/>
CC or send an email to license@isb-sb.ch).
CC -----
CC EMBL: M61095; AAA63423.1; -
CC Melaria; Receptor; Glycoprotein; Signal; Transmembrane.
CC
CC SIGNAL 1 20 POTENTIAL.

```

FT CHAIN 21 1070 DUFFY RECEPTOR.
FT DOMAIN 21 1007 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1008 1025 POTENTIAL.
FT DOMAIN 1026 1070 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 183 183 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 255 255 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 351 351 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 420 420 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 715 715 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 787 787 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 825 825 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 903 903 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 938 938 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1070 AA: 119683 MW: CB051DF13E294603 CRC64;

```

Query Match 5.9%; Score 134.5; DB 1; Length 1070;

Best Local Similarity 20.0%; Pred. No. 0.88; Mismatches 220; Indels 89; Gaps 21;

```

Matches 97; Conservative 80;
QY 2 NKEYSISKENTHNGLRMSIVPKSTGKGFEDGDAISKLHDEEYGAALIRNNMOHIMG 61
Db 455 NKFTSVKNAEKVQTAGIVTPY-----DILKQ--ELDEFNEVAFENEINKRDG 499
QY 62 PVTAIKLLG-EENKENTPRNLTIIIPASMTAKAHSKDKKRPORDQAQSPVSKSTHR 120
Db 500 AYIELCVSEAEARKNTQEVVTNVDMAKSAQ-TNSNPISQPVDSKAKEVP--GDSTH- 555
QY 121 IOHNDYLLKHLKV-KKIPSDFECSGYTDLO---ERGDNDISPSGQGPKEIDRGKE- 175
Db 556 --GAVNSGQDSSTGTGKAVTGGONGNOTPAESDVQSRSDIAESVAKKNDPOKSVKRSDD 613
QY 176 ---ATGPDLEGKDIQTGPAGPSEASTHLDTKRPG-----YNEI 211
Db 614 TASYTGIAEAGKE-NLGASNSRPEST-VEANSPGDDTVNSASIPVVGSENPVLTPTNGL 671
QY 212 PERENGNGN-----TIGTRDETAKEDAVSVLEGSNDIMGSTNFKELPREGN 261
Db 672 RHSGKNDSDSDGPAESMANPDNSNSKGETGKGQDNDMAKATKDDSSNSDGTG----SATGD 726
QY 262 RVDAGSONAHQGVFEHYPPAPSKERKREGSS--DAAESTNYNEIPKNGKST---RKG 315
Db 727 TTDAVDREINKGVPEDEKDTYVSGKDGGEEDNSANKDAIYVGEDRIRENSAGSTNDRSK 786
QY 316 VDHSSNRQATLINEQR--FPSKGSQGLPIPSRG--LDNEIKNEMDSFNGPSHENITTHG 371
Db 787 NDETEKNGASTPDSKSEDAATALSKTESLESTESGDRTTNDTNSLENKNGKEDLQKHD 846
QY 372 RKYHYVPHRQNNSTR-----NKGMPQCKGSMGRQPHSNR-RFSSRRDDSSSESDSG 422
Db 847 FKSNDTPNEEPNSDQTTDAEGHDROSTIKNDKAE--RRKHMKDFTFKNTNSHHLNSNNNL 904
QY 423 SSSESDD 428
Db 905 SNGKLD 910

```

Search completed: April 11, 2003, 12:02:36
Job time : 22 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2003, 11:59:32 ; Search time 34 Seconds

(Without alignments)
2605.890 Million cell updates/sec

Title: US-09-700-696B-2

Perfect score: 2279

Sequence: 1 VKKEYSISNKNENTHNGLRMS.....RRDSSSESDSGSSSESDG 430

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08

Maximum Match 1008

Listing first 45 summaries

Database :

1: SP archaea: *
2: SP bacteria: *
3: SP fungi: *
4: SP human: *
5: SP_invertebrate: *
6: SP_mammal: *
7: SP_mhc: *
8: SP_Organelle: *
9: SP_phage: *
10: SP_plant: *
11: SP_rodent: *
12: SP_virus: *
13: SP_vertebrate: *
14: SP_unclassified: *
15: SP_virus: *
16: SP_bacteriap: *
17: SP_archaeap: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2276	99.9	525	4 Q9NO76	Q9NO76 homo sapien
2	2058.5	90.3	500	6 Q95K65	Q95K65 macaca fasc
3	2058.5	90.3	555	6 Q9NO76	Q9NO76 macaca fasc
4	2054.5	90.1	555	6 Q9GM13	Q9GM13 macaca fasc
5	983.5	43.2	433	11 Q92411	Q92411 mus musculu
6	938.5	41.2	433	11 Q9ES02	Q9ES02 rattus norv
7	644	28.3	306	6 Q9BE56	Q9BE56 macaca fasc
8	182	8.0	1695	5 Q9BK62	Q9BK62 hydra atten
9	165	7.2	1895	5 Q9VW55	Q9VW55 drosophila
10	157	6.9	970	11 Q9VBY1	Q9VBY1 rattus norv
11	156.5	6.9	2615	16 Q9CP11	Q9CP11 pasteurilla
12	151	6.6	847	6 Q95K77	Q95K77 canis famli
13	151	6.6	1422	6 Q95K04	Q95K04 canis famli
14	150	6.6	2829	13 P70039	P70039 xenopus lae
15	150	6.6	3060	5 Q9VAV4	Q9VAV4 drosophila
16	149	6.5	687	4 Q9H4U2	Q9H4U2 homo sapien

17	148	6.5	2081	10 Q9LH98	Q9LH98 arabidopsis
18	147	6.5	1461	5 Q95K59	Q95K59 plasmodium
19	146	6.4	485	2 Q9AH49	Q9AH49 neisseria m
20	146	6.4	2174	5 Q9GCR0	Q9GCR0 drosophila
21	144.5	6.3	665	5 Q9G229	Q9G229 plasmodium
22	143.5	6.3	1207	5 Q21535	Q21535 caenorhabdi
23	143	6.3	791	13 Q9DGL1	Q9DGL1 figu rubrip
24	142.5	6.3	635	10 Q40363	Q40363 medicago sa
25	142	6.2	465	4 Q03838	Q03838 homo sapien
26	142	6.2	1445	3 Q74229	Q74229 kluyveromyc
27	141.5	6.2	381	10 Q9FK45	Q9FK45 arabidopsis
28	141.5	6.2	1579	13 Q90X36	Q90X36 xenopus lae
29	140.5	6.2	990	4 Q15206	Q15206 homo sapien
30	140.5	6.2	1218	4 Q05331	Q05331 homo sapien
31	140	6.1	852	10 Q9S255	Q9S255 homo sapien
32	140	6.1	772	4 Q14241	Q14241 homo sapien
33	139.5	6.1	1035	5 Q21380	Q21380 caenorhabdi
34	138.5	6.1	848	5 Q94419	Q94419 schizosacch
35	138	6.1	599	5 Q74354	Q74354 schizosacch
36	138	6.1	931	5 Q9VF92	Q9VF92 drosophila
37	138	6.1	2897	5 Q9VID9	Q9VID9 drosophila
38	137.5	6.0	798	4 Q9H4U3	Q9H4U3 homo sapien
39	137.5	6.0	3111	5 Q9VH10	Q9VH10 drosophila
40	137	6.0	611	10 Q41042	Q41042 pisum sativ
41	137	6.0	1125	16 Q98P09	Q98P09 mycoplasma
42	137	6.0	2112	5 Q9VEL9	Q9VEL9 drosophila
43	136.5	6.0	585	10 Q9M255	Q9M255 arabidopsis
44	136.5	6.0	1231	5 Q9X221	Q9X221 drosophila
45	135.5	5.9	1211	11 Q35233	Q35233 mus musculu

ALIGNMENTS

RESULT 1

ID	Q9NO76	PRELIMINARY:	PRT:	525 AA.
AC	Q9NO76:			
DT	01-OCT-2000 (TREMURel. 15, Created)			
DT	01-OCT-2000 (TREMURel. 15, Last sequence update)			
DT	01-DEC-2001 (TREMURel. 19, Last annotation update)			
DE	Matrix extracellular phosphoglycoprotein precursor.			
GN	MEPE.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=BONE;			
RX	MEDLINE-2039567; PubMed-10945470;			
RA	Rowe P.S.N., De Zoyza P.A., Dong R., Wang H.R., White K.E.,			
RA	Reons M.J., Oudet C.L.;			
RT	"MEPE, a new gene expressed in bone marrow and tumors causing			
RT	osteomalacia.";			
RL	Genomics 67:54-68(2000).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE-21309068; PubMed-11414762;			
RA	Argiro L., Desparats M., Glorieux F.H., Ecartot B.;			
RT	"Mepe, the gene encoding a tumor-secreted protein in oncogenic			
RT	hypophosphatemic osteomalacia, is expressed in bone.";			
RL	Genomics 74:342-351(2001).			
DR	EMBL; AJ276396; CAB97250.1; -			
DR	EMBL; AF325916; AAK70343.1; -			
KW	SIGNAL.			
FT	SIGNAL.			
FT	CHAIN			
SO	SEQUENCE			
	525 AA; 58419 MW; 0977CA6E871CA9E5 CRC64;			
	POTENTIAL.			
	MATRIX EXTRACELLULAR PHOSPHOGLYCOPROTEIN.			
	Query Match			
	Best Local Similarity			
	Matches			
	429; Conservative			
	1; Mismatches			
	0; Gaps			
	0;			

```

OY 1 VNKEYSISNKNENTHNGLRMSIYKSTGKGFEDGDDAISKLDHOEEYGAALIRNMOMHIM 60
D 1 LNKEYSISNKNENTHNGLRMSIYKSTGKGFEDGDDAISKLDHOEEYGAALIRNMOMHIM 155
OY 61 GPTAATKILGEEKENTPRNVLNITIPASMYAKAHSKDKKRPORSQAOKSPVKSSTHR 120
D 156 GPTAATKILGEEKENTPRNVLNITIPASMYAKAHSKDKKRPORSQAOKSPVKSSTHR 215
OY 121 IOHNDYKHLKLSKYKKIPSDFEESGYTDLQERGNDISPSSGDDQPFKDIIPGKEATGPD 180
D 216 IOHNDYKHLKLSKYKKIPSDFEESGYTDLQERGNDISPSSGDDQPFKDIIPGKEATGPD 275
OY 181 LEGKDIOGTFAPSPAESESTHLDTKKPGYNEIPEREENGNTIGTRDETAKEADAVDSL 240
D 276 LEGKDIOGTFAPSPAESESTHLDTKKPGYNEIPEREENGNTIGTRDETAKEADAVDSL 335
OY 241 EGSNDIMGSTNFKELPGREGNRYDAGSONAHQKVEFYHPPAPSKERKESSDAESTN 300
D 336 EGSNDIMGSTNFKELPGREGNRYDAGSONAHQKVEFYHPPAPSKERKESSDAESTN 395
OY 301 YNEIPRNGKSTRGVDSHNRQATLNEKORPPSKSGGLPIPSRGLDNETIKNEMDSFN 360
D 396 YNEIPRNGKSTRGVDSHNRQATLNEKORPPSKSGGLPIPSRGLDNETIKNEMDSFN 455
OY 361 GSHENIITHGKRKYHYVPHRONNSTRNKGMPOGKSGMGRPHSNRFRSSRRRDESSSD 420
D 456 GSHENIITHGKRKYHYVPHRONNSTRNKGMPOGKSGMGRPHSNRFRSSRRRDESSSD 515
OY 421 SSSSESDDG 430
D 516 SSSSESDDG 525

```

RESULT 2

095KG5 PRELIMINARY; PRT: 500 AA.

```

AC 095KG5;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Hypothetical 55.6 kDa protein.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
OC NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TEMPORAL LOBE, RIGHT;
RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
RA Suzuki Y., Sugano S., Hashimoto K.;
RT "Isolation of full-length cDNA clones from macaque brain cDNA
RT libraries.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB060891; BAB46894.1;
KW Hypothetical protein.
SQ SEQUENCE 500 AA; 35577 MW; 918D265AD8EDC7BC CRC64;

```

Query Match 90.3%; Score 2058.5; DB 6; Length 500;
 Best Local Similarity 90.9%; Pred. No. 9.3e-120;
 Matches 391; Conservative 13; Mismatches 25; Indels 1; Gaps 1;

```

OY 1 VNKEYSISNKNENTHNGLRMSIYKSTGKGFEDGDDAISKLDHOEEYGAALIRNMOMHIM 60
D 72 LNKEYSISNKNENTHNGLRMSIYKSTGKGFEDGDDAISKLDHOEEYGAALIRNMOMHIM 131
OY 61 GPTAATKILGEEKENTPRNVLNITIPASMYAKAHSKDKKRPORSQAOKSPVKSSTHR 120
D 132 GPTAATKILGEEKENTPRNVLNITIPASMYAKAHSKDKKRPORSQAOKSPVKSSTHR 191
OY 121 IOHNDYKHLKLSKYKKIPSDFEESGYTDLQERGNDISPSSGDDQPFKDIIPGKEATGPD 180
D 121 IOHNDYKHLKLSKYKKIPSDFEESGYTDLQERGNDISPSSGDDQPFKDIIPGKEATGPD 180

```

```

D 192 IOHNDYKHLKLSKYKKIPSDFEESGYTDLQERGNDISPSSGDDQPFKDIIPGKEATGPD 251
OY 181 LEGKDIOGTFAPSPAESESTHLDTKKPGYNEIPEREENGNTIGTRDETAKEADAVDSL 240
D 252 LEGKDIOGTFAPSPAESESTHLDTKKPGYNEIPEREENGNTIGTRDETAKEADAVDSL 311
OY 241 EGSNDIMGSTNFKELPGREGNRYDAGSONAHQKVEFYHPPAPSKERKESSDAESTN 300
D 312 EGSNDIMGSTNFKELPGREGNRYDAGSONAHQKVEFYHPPAPSKERKESSDAESTN 371
OY 301 YNEIPRNGKSTRGVDSHNRQATLNEKORPPSKSGGLPIPSRGLDNETIKNEMDSFN 360
D 372 YNEIPRNGKSTRGVDSHNRQATLNEKORPPSKSGGLPIPSRGLDNETIKNEMDSFN 431
OY 361 GSHENIITHGKRKYHYVPHRONNSTRNKGMPOGKSGMGRPHSNRFRSSRRRDESSSD 420
D 432 GSHENIITHGKRKYHYVPHRONNSTRNKGMPOGKSGMGRPHSNRFRSSRRRDESSSD 490
OY 421 SSSSESDDG 430
D 491 SSSSESDDG 500

```

RESULT 3

09N076 PRELIMINARY; PRT: 555 AA.

```

AC 09N076;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE Matrix extracellular phosphoglycoprotein.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
OC NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TEMPORAL LOBE, RIGHT;
RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
RA Suzuki Y., Sugano S., Hashimoto K.;
RT "Isolation of full-length cDNA clones from macaque brain cDNA
RT libraries.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB046056; BAB01638.1;
SQ SEQUENCE 555 AA; 61979 MW; CCFELA99ADA19E4 CRC64;

```

Query Match 90.3%; Score 2058.5; DB 6; Length 555;
 Best Local Similarity 90.9%; Pred. No. 1.1e-119;
 Matches 391; Conservative 13; Mismatches 25; Indels 1; Gaps 1;

```

OY 1 VNKEYSISNKNENTHNGLRMSIYKSTGKGFEDGDDAISKLDHOEEYGAALIRNMOMHIM 60
D 127 LNKEYSISNKNENTHNGLRMSIYKSTGKGFEDGDDAISKLDHOEEYGAALIRNMOMHIM 186
OY 61 GPTAATKILGEEKENTPRNVLNITIPASMYAKAHSKDKKRPORSQAOKSPVKSSTHR 120
D 187 GPTAATKILGEEKENTPRNVLNITIPASMYAKAHSKDKKRPORSQAOKSPVKSSTHR 246
OY 121 IOHNDYKHLKLSKYKKIPSDFEESGYTDLQERGNDISPSSGDDQPFKDIIPGKEATGPD 180
D 247 IOHNDYKHLKLSKYKKIPSDFEESGYTDLQERGNDISPSSGDDQPFKDIIPGKEATGPD 306
OY 181 LEGKDIOGTFAPSPAESESTHLDTKKPGYNEIPEREENGNTIGTRDETAKEADAVDSL 240
D 307 LEGKDIOGTFAPSPAESESTHLDTKKPGYNEIPEREENGNTIGTRDETAKEADAVDSL 366
OY 241 EGSNDIMGSTNFKELPGREGNRYDAGSONAHQKVEFYHPPAPSKERKESSDAESTN 300
D 367 EGSNDIMGSTNFKELPGREGNRYDAGSONAHQKVEFYHPPAPSKERKESSDAESTN 426
OY 301 YNEIPRNGKSTRGVDSHNRQATLNEKORPPSKSGGLPIPSRGLDNETIKNEMDSFN 360
D 427 YNEIPRNGKSTRGVDSHNRQATLNEKORPPSKSGGLPIPSRGLDNETIKNEMDSFN 486

```

QY 361 GSHENITTHGKRYHYVPHRONNSTRNKGMPGKSGMRPHSNRRSSRRDSSSSD 420
 DB 487 GGSNNEN-IPHSKRYHYVPHRONNSTRNKGMPGKSGMRPHSNRRSSRRDSSSSD 545
 QY 421 SSSSSSSDGD 430
 DB 546 SSSSSSSDGD 555

RESULT 4

Q9GM13 PRELIMINARY; PRT: 555 AA.
 AC Q9GM13;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
 DE MEPE protein.
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecoidea; Macaca.
 NCBI_TaxID=9541;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain PARIETAL LOBE;
 RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
 RA Suzuki Y., Sugano S., Hashimoto K.;
 RT "Isolation of full-length cDNA clones from macaque brain cDNA
 RT libraries."
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB050259; BAB17010.1; -;
 SO SEQUENCE 555 AA; 61950 MW; BD4D47E8803BA9E2 CRC64;

Query Match 90.1%; Score 2054.5; DB 6; Length 555;
 Best Local Similarity 90.7%; Pred. No. 1.9e-119;
 Matches 390; Conservative 14; Mismatches 25; Indels 1; Gaps 1;

QY 1 VNEVSTSNKENTHNGLRMSIYKSTGKGFEGDGDALISLHDOEEGALLRNNOHITM 60
 DB 127 LNEVSTSNKENTHNGLRMSIYKSTGKGFEGDGDALISLHDOEEGALLRNNOHITM 186
 QY 61 GPVTAIKLLEENKENTPRVNLITIPASMYAKAHSKDKKPPORDSOAKSPVKSSTR 120
 DB 187 GPVTAIKLLEENKOSKPKVNLKIPASMYAKAHSKDKKPPORDSOAKSPVKSSTR 246
 QY 121 IOHNITVYKHLKSKYKIPSPFEGSGYTDLOERGDNDISPFSGGQPFKDIPIKGEATGPD 180
 DB 247 TQHNIDYKHLKSKYKIPSPFEGSGYTDLOERGDNDISPFSGGQPFKDIPIKGEATGPD 306
 QY 181 LECKDIQTGFAGPSEASTHLDTKKPGYNEIPEREENGNTIGTRDTAKKADAVDSLY 240
 DB 307 LECKDIQTGFAGPSEASTHLDTKKPGYNEIPEREENGNTIGTRDTAKKADAVDSLY 366
 QY 241 EGSNDIMGSTNKEKELPREGGNRVDAQSNAHOGKVEFHYPPAPSKERKREGSSDAESTNY 300
 DB 367 EGSNDIMGSTNKEKELPREGGNRVDAQSNAHOGKVEFHYPPAPSKERKREGSSDAESTNY 426
 QY 301 YNEIPANGKSTRKGVDSNRNQAATLNEKORFPSPKSGSLGPIPSRGLDNETKNDMSFN 360
 DB 427 YNEIPANGKSTRKGVDSNRNQAATLNEKORFPSPKSGSLGPIPSRGLDNETKNDMSFN 486
 QY 361 GSHENITTHGKRYHYVPHRONNSTRNKGMPGKSGMRPHSNRRSSRRDSSSSD 420
 DB 487 GGSNNEN-IPHSKRYHYVPHRONNSTRNKGMPGKSGMRPHSNRRSSRRDSSSSD 545
 QY 421 SSSSSSSDGD 430
 DB 546 SSSSSSSDGD 555

RESULT 5
 Q92411

ID Q92411 PRELIMINARY; PRT: 433 AA.
 AC Q92411;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Matrix extracellular phosphoglycoprotein precursor.
 GN MEPE.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J;
 RA MEDLINE=21309068; PubMed=11414762;
 RA Argiro L., Desbarats M., Glorieux F.H., Escart B.;
 RT "Mepe, the gene encoding a tumor-secreted protein in oncogenic
 RT hypophosphatemic osteomalacia, is expressed in bone."
 RL Genomics 74:342-351(2001).
 DR EMBL: AF314964; AAK70342.1; -;
 DR MGD: MGI:2137384; Mepe.
 KW Signal.
 FT SIGNAL.
 FT CHAIN 19 18 POTENTIAL.
 SO SEQUENCE 433 AA; 45984 MW; 7CD603CA8CB41B0 CRC64;

Query Match 43.2%; Score 983.5; DB 11; Length 433;
 Best Local Similarity 50.5%; Pred. No. 2.7e-53;
 Matches 217; Conservative 52; Mismatches 134; Indels 27; Gaps 8;

QY 6 SISNKEENTHNGLRMSIYKSTGKGFEGDGDALISLHDOEEGALLRNNOHITM 65
 DB 26 SCGNODSTHDKDLAASYDPDYDTEGTEGQGLLHPPEODRYGALLRNITQPVKSLVTG 85
 QY 66 IYLLLEENKENTPRVNLITIPASMYAKAHSKDKKPPORDSOAKSPVKSSTR 122
 DB 86 AELRREGNOEKRPQSVLVIYPADYNDKAVSLKDIKNGSYLLTQSSPVKSKHTKHTROT 145
 QY 123 NHIDLKHLKSKYKIPSPFEGSGYTDLOERGDNDISPFSGGQPFKDIPIKGEATGPD 181
 DB 146 RSTHILTHLPQIKTIPSPDEGSGPDLVLRKDNVPPSPGQGHMHPGKGGAGSGPE 204
 QY 182 ECKDIQTGFAGPSEASTHLDTKKPGYNEIPEREENGNTIGTRDTAKKADAVDSLY 241
 DB 205 --SSTRPLSGSSKAEVIDIPHMSGLSNEIPREGHGSAATADKAAQAGSGAGSLVG 262
 QY 242 GSNIDIMGSTNKEKELPREGGNRVDAQSNAHOGKVEFHYPPAPSKERKREGSSDAESTNY 301
 DB 263 GSNIDIMGSTNKEKELPREGGNRVDAQSNAHOGKVEFHYPPAPSKERKREGSSDAESTNY 321
 QY 302 NEIPANGKSTRKGVDSNRNQAATLNEKORFPSPKSGSLGPIPSRGLDNETKNDMSFN 361
 DB 322 NEIPANGKSTRKGVDSNRNQAATLNEKORFPSPKSGSLGPIPSRGLDNETKNDMSFN 375
 QY 362 PSHEMITTHGKRYHYVPHRONNSTRNKGMPGKSGMRPHSNRRSSRRDSSSSD 420
 DB 376 --EN-----HYVFGHGNLITPKNKMGSRGSMPSRRNSHRASTRQR-DSSSSS 423
 QY 421 SSSSSSSDGD 430
 DB 424 SSSSSSSDGD 433

RESULT 6
 Q9ES02 PRELIMINARY; PRT: 435 AA.

AC Q9ES02;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE Osteoregulin.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SPRAGUE-DAWLEY;
 RX MEDLINE=20549633; PubMed=10967096;
 RA Petersen D.N., Tkalecic G.T., Mansolf A.L., Rivera-Gonzalez R.,
 Brown T.A.;
 RT "Identification of Osteoblast/Osteocyte Factor 45 (OF45), a Bone-
 specific cDNA Encoding an RGD-containing Protein That Is Highly
 Expressed in Osteoblasts and Osteocytes";
 RL J. Biol. Chem. 275:36172-36180(2000).
 DR EMBL: AF260922; AAG3336.1; -
 SQ SEQUENCE 435 AA; 46515 MW; D587F82968426BCB CRC64;
 Query Match 41.2%; Score 938.5; DB 11; Length 435;
 Best Local Similarity 49.4%; Pred. No. 1.7e-50;
 Matches 211; Conservative 45; Mismatches 146; Indels 25; Gaps 8;
 OY 9 NKENTHGLRMSIYPKSTGNKGFEDGDAISKLDQEEYGAALIRNNMOMHMGVTAIKL 68
 DB 29 NQGNH---LASVKEPEWVGKTEGGRDAPLHLDONKOGATLLRNITQPKSLVTGTEV 85
 OY 69 LGEENKENTPRNVNIIIPASNNYAKAHSKDKKPPQDSQAOKSPVSKSTHRIQNDIYL 128
 DB 86 QSDNRKKEPKQSVLSTPTDVHNTNDYSEDTENQDRDLLONSFGQSKHPRARSTHYL 145
 OY 129 KHLKSVKRPDFFSGYTDLOERGDNDISFSGDGGPFKDIPEGKATGPDLE---GKD 185
 DB 146 THLPQIRKLSDFFEDSASPDLLVGRGNDVPPFSGDGGHFMHTPDRCGAVGSDPESSGHP 205
 OY 186 IOTGAGAGSEASHLDPKKRGVNEIPREENGNTIGTRETKEADAVVSLVEGSDND 245
 DB 206 V---SGSSNVEIYDPHTNGLSNTEPGECHIGATVTRKKTLOGAGSADVSLVEGSE 261
 OY 246 IMGSTNFKELPGREGNRYDAGSQNAHQGVFHYPPAPSKRREKGSDDAAESTNYNEIP 305
 DB 262 ITGSKTFELGKEGNRRDASQNAHQGVFHYPPAPSKRREKGSDDAAESTNYNEIP 321
 OY 306 KNGKSTKGVDSHNRNOATLEKORFSPKSKSGQLPSPSGLDNEIKENMDFNGPSHE 365
 DB 322 KSSKSGAKDEESKGNVTLTESGRFPKGGQS---SHSLGVEVSEEDSSLSLRE 377
 OY 366 NT-ITHGKRYHVPRQNNSTFNKCMPOGKGSW-GROPHSNRRFSRRRSDSSSDSGS 423
 DB 378 GTAIHAKRTSH-----PTNRGMSQRRGSMARRHPRHRYSTQR-DSSSSSSGS 428
 OY 424 SSESDDG 430
 DB 429 SSSSGD 435
 RESULT 7
 OY9B56 PRELIMINARY; PRT; 306 AA.
 AC OY9B56;
 DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE Hypothetical 35.0 kDa protein.
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
 OC Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecoidea; Macaca.
 OX NCBI_TaxID=9541;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=FRONTAL LOBE LEFT;
 RA Osaka N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
 Suzuki Y., Sugano S., Hashimoto K.;
 RT "Isolation of full-length cDNA clones from macaque brain cDNA
 libraries";
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL: AB056814; BAB39339.1;
 KW Hypothetical protein.
 SQ SEQUENCE 306 AA; 35013 MW; 1903217AFD05B083 CRC64;
 Query Match 28.3%; Score 644; DB 6; Length 306;
 Best Local Similarity 89.9%; Pred. No. 1.8e-32;
 Matches 124; Conservative 4; Mismatches 10; Indels 0; Gaps 0;
 OY 1 YNKEYSISKNENHNLMSIYPKSTGNKGFEDGDAISKLDQEEYGAALIRNNMOMH 60
 DB 127 LNKESYISKNENHNLMSIYPKSTGNKGFEDGDAISKLDQEEYGAALIRNNMOMH 186
 OY 61 GPVTAIKLGEENKENTPRNVNIIIPASNNYAKAHSKDKKPPQDSQAOKSPVSKSTH 120
 DB 187 GPVTAIKLGEENKQSKPKVNLNKPASNNYAKAHSKDKKPPQDSQAOKSPVSKSTH 246
 OY 121 IOHNDIYKHLKSVKRP 138
 DB 247 TQHNIDYKHLKSVKRSK 264

RESULT 8
 OY9BKL2 PRELIMINARY; PRT; 1695 AA.
 AC OY9BKL2;
 DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE Tight junction protein ZO-1.
 GN ZO-1.
 OS Hydra attenuata (Hydra) (Hydra vulgaris).
 OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroidea; Anthomedusae;
 OC Hydraeae; Hydra.
 OX NCBI_TaxID=6087;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=21025768; PubMed=11225567;
 RA Fel K., Yan L., Zhang J., Sarraz M.P. Jr.;
 RT "Molecular and biological characterization of a zonula occludens-1
 homologue in Hydra vulgaris, named HZO-1";
 RL dev. Genes Evol. 210:611-616(2000).
 DR EMBL: AF230482; AKR28322.1; -
 DR HSP: P29476; 10AV.
 DR InterPro: IPR000619; Guanylate_kin.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001452; SH3.
 DR InterPro: IPR000906; ZU5.
 DR Pfam: PF00625; Guanylate_kin; 1.
 DR Pfam: PF00595; PDZ; 3.
 DR Pfam: PF00791; ZU5; 1.
 DR SMART: SM00072; GuKc; 1.
 DR SMART: SM00228; PDZ; 3.
 DR SMART: SM00326; SH3; 1.
 DR SMART: SM00218; ZU5; 1.
 DR PROSITE: PS50052; GUANYLATE_KINASE_2; 1;
 DR PROSITE: PS50106; PDZ; 3.
 DR PROSITE: PS50002; SH3; 1.
 DR SEQUENCE 1695 AA; 190878 MW; AB0E7D09007316F1 CRC64;
 Query Match 8.0%; Score 182; DB 5; Length 1695;
 Best Local Similarity 23.1%; Pred. No. 0.0062;
 Matches 110; Conservative 72; Mismatches 175; Indels 120; Gaps 23;
 OY 10 KENTHGLRMSIYPKSTG-NKGFEDGDAI-----SKLDQEEYGAALIRNNMOMH 59
 DB 25 KKSARQGFIAI---SGLDNPHFTGTSTIVSDIYVSGSPADGKLKGDILISVNERNV 81
 OY 60 MG-----PYTAIKLGEENK-----NTb-RNVNIIIPASNNYAK-----AHSKDKKP 102
 DB 82 DRSRHDAEALKAAGMEARMEIKRPSINPKKNDINDNSVNGSKETDSEGRGSRKP 141
 OY 103 QDSQAOKSPVSKSTHRIQ---HNIDYKHLKSVKRIISDFEGSGYT---DLOERGDNDI 157
 DB 103 QDSQAOKSPVSKSTHRIQ---HNIDYKHLKSVKRIISDFEGSGYT---DLOERGDNDI 157

```

Db 142 K--EMEQSGKKHSHSHOKEDHDSEKPRHRSKREKEGEKSSRSNKLSDNADNR 199
Oy 158 SPFGDGPFPKNDPGKGEATGPDLEKDIQGFAGPSEASTHLDTKKPGYNEIPEEBN 217
Db 200 SRSRHSR-----SKNREDSGCD-----KSEKSHRSR-----NRDN 234
Oy 218 GGTGTGTRD-----EFAKADAVDVLVEGSNDI-MGSTNFEELGREGNRYDAGS- 267
Db 235 AEGESHSRHSRKNREDAETGKSEKSHRSRKNREDAETGKSEKSHRSRKNREDAETG 294
Oy 268 QNAHOGKVEFHYPPAP-----SEKRRKSGSSDAESTYNEIPIKNGKSTRGVDHNS- 320
Db 295 EKSHRSRKNRDEPTEEEKSLRSKSTREDLKTTESEKSKIPKSGKASLGMEKENS 354
Oy 321 -----RNOATLNEKORPPSKGKSOGDPIPSRGDNEIKNEMDSFNGSHENITTHGK 373
Db 355 DRKSKTRREDAGEEERKARIMS--TOSKP-----NITESSDK 388
Oy 374 YHVPVRONNSTR--NKGPDOGK--SMGROPHSNRRSSRRRDDSESSSDGSSSES 427
Db 389 -----HRSRSTREDTEGVEKPRSRSTORKPMDMEKLSRANDEGKASRSRKS 440

RESULT 9
OyVMS5 PRELIMINARY: PRT: 1895 AA.
AC 09VMS5;
Dt 01-MAY-2000 (TREMBlrel. 13, Created)
Dt 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
Dt 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE CG15040 protein.
GN CG15040.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abail J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Bencos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borkov A., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA De Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Goddek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.J.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hosteln D., Houston K.A., Howland T.J., Wei M.-H., Idegami C.,
RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kemison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodali C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., Meperson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshneft A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pauley J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheefer F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier B., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,

```

```

RA Williams S.M., Woodage T., Morley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL: AE003509; AAF4863.1;
DR FlyBase: FBgn0030940; CG15040.
DR InterPro: IPR002965; P-rich extensn.
DR PRINTS: PR01217; PRICHEXTNSN.
SQ SEQUENCE 1895 AA; 208652 MW; DB8540F2EC4A1CA8 CAC64;

Query Match 7.2%; Score 165; DB 5; Length 1895;
Best Local Similarity 18.6%; Pred. No 0.08;
Matches 74; Conservative 74; Mismatches 168; Indels 82; Gaps 16;

Oy 9 NKENTHGLRMSIYPKSGNKGFEEDGDAISKLDDQEGYGAALLRNMOMHMGVTAIKL 68
Db 1557 NENSHNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1593
Oy 69 LGEENKENTPRNVLIITIPASMYAKAH-----SKDKKRPQDSQAOKSPVSKSTHR 120
Db 1594 --SKNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1646
Oy 121 IOHNDLYLKLHLSKYKIPDSFEGSGYTDLOERGNDISPFSGDGPFPKNDPGKGEATGPD 180
Db 1647 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1696
Oy 181 LEGKTIQGFAGPSEASTHLDTKKPG--YNEIPEEBNGG-NTIGTDETAKEADAVDV 237
Db 1697 FQSKN-QNNNIOSKNODNYYNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1751
Oy 238 SLVEGSNDI-MGSTNFEELGREGNRYDAGSQAHAOGKVEFHYPPAPSEKRRKSGSDA 296
Db 1752 ---QNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1799
Oy 297 ESTNTNIEIPKNGKSTRGVDHNSRNOATLNEKORPPSKGKSOGDPIPSRGDNEIKNEM 356
Db 1800 QNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1856
Oy 357 DSENGPSHENITTHGKRYHYVPHRONNSTRNKGPOGK 394
Db 1857 QSKNNKNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1884

RESULT 10
OyVMS5 PRELIMINARY: PRT: 970 AA.
AC 08VBY1;
Dt 01-MAR-2002 (TREMBlrel. 20, Created)
Dt 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
Dt 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Phosphophoryn, precursor (Dentin stallophosphoprotein) (Dentin
DE stallophosphoprotein precursor) (stallophosphoprotein-phosphophoryn
DE DSP-pp53).
GN DSP-pp OR DSP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY;
RX MEDLINE=21450300; PubMed=11566357;
RA Ritchie H.H., Wang L.H., Knudsen K.;
RT "A rat DSP-pp53 transcript encodes for a 523 amino acid pp protein
and its genomic organization."
RL Submitted (MAY-2000) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY;
RX MEDLINE=21450300; PubMed=11566357;
RA Ritchie H.H., Wang L.H., Knudsen K.;
RT "A novel rat 523 amino acid phosphophoryn: nucleotide sequence and
genomic organization."

```

RESULT	11
09Cpil	
ID	09Cpil
AC	09Cpil
DT	01-JUN-2001 (TREMblrel. 17, Created)
DT	01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT	01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE	PFMB1.
GN	PFMB1 OR PM0057.
OS	Pasteurella multocida.
OC	Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC	Pasteurella.
OX	NCBI_TaxID=747;

	RESULT	12
	095KP7	
ID	095KP7	PRELIMINARY;
AC	095KP7	PRT; 847 AA.
DT	01-DEC-2001 (TREMBLE)	19, Created
DT	01-DEC-2001 (TREMBLE)	19, last sequence update)
DT	01-JUN-2002 (TREMBLE)	21, last annotation update)
DE	Treacle (Fragment).	
GN	TCOFL.	
OS	Canis familiaris (Dog).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.	
OX	NBI_TaxID=9615;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RA	Haworth K.E.;	
RT	"Canine TCOFL: Cloning, chromosome assignment and genetic analysis in	
RT	dogs with different face and head types.";	
LT	Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.	

DR EMBL: AJ302713; CAC88122.1; JOINED.
 DR EMBL: AJ302714; CAC88122.1; JOINED.
 DR EMBL: AJ302715; CAC88122.1; JOINED.
 DR EMBL: AJ302716; CAC88122.1; JOINED.
 DR EMBL: AJ302717; CAC88122.1; JOINED.
 DR EMBL: AJ302718; CAC88122.1; JOINED.
 DR EMBL: AJ302719; CAC88122.1; JOINED.
 DR InterPro: IPR003993; treeacle.
 DR Pfam: PF03546; treeacle; 3.
 DR NON_TER 847
 FT SEQUENCE 847 AA; 85646 MW; C5A763FDF1CA2235 CRC64;

Query Match 6.6%; Score 151; DB 6; Length 847;
 Best Local Similarity 21.5%; Pred. No. 0.22; Mismatches 171; Indels 118; Gaps 19;
 Matches 96; Conservative 61;

12 NTHNGLRMSIYKSTGNKGFEDGDAISKLDQEEYGALLIRNMQHIM-----GP 62
 109 STNSSVPGVLPSTSTKEGVAKTNKA-SKVNSTPHPAS--AKAVAHILSGRSPRKSAGP 165
 63 VTAIKLGEENKENTPRNVNLITPASMYAKAHKDKKKPORDSOAQSPYKSKSTHRIQ 122
 166 SANITLIVSETEEGS-----VPALGTTAKPGMASAN--QADSSSETS----- 205
 123 HNIDYLKHLSTKVKKIPSEEGSGYTDLOERGDNDISPESGDQPFKIDIPKGEATGPDL 182
 206 -----SSDETIVEYKASEKIYQAKAAGPYKGTGPGATPAP--- 243
 183 GKDIOTGFAGPSEASTHLDTKPKPYNEIPEREENGNTIGTRDETAKADAVSYLVEG 242
 244 -----PGKAGPSAAQ---KTEKP-----KEDS-----DSSEDSDEE----- 274
 243 SNDIMGSTNFKELPGREGNRVDAGSONAHQ-----GKVEFHYPPAPSKERKEG 291
 275 -EPPAKTLPQVKGSGKTPQVKAASAKESPRKGVPPVPGKV-----GPAAGAKKGAG 329
 292 SSDAESTNYNEIPKNGKS---TRKGVDSNRNOATLNEKORFPPSKGSQ-----GLPI 343
 330 EEDPDSTSEDSSEEAAPTAVPTSPYQAKPSGQNSQVRTASGPKVGPQAKGAPATPV 389
 344 PSRGIDNEIKNEDS-FNGSHENITTHGRKHYVPHRONNSTRNKMGPOGKSGWRGPH 402
 390 GKOEDESSSEESSEEGGAAPQAKSGK-----IPQVRAASGPAKGPQKAG-----PA 440
 403 SNRRFSRRDDSSSDSGSSSESD 428
 441 ATQAKAKMSKD-SSESSESESEEE 465

RESULT 13
 095K04 PRELIMINARY; PRT; 1422 AA.
 AC 095K04; (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 21, Last annotation update)
 DE treeacle.
 GN TCOF1.
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OC NCBI_TaxID=9615;
 RP SEQUENCE FROM N.A.
 RA Haworth K.E., Islam I., Breen M., Putt W., Binns M., Hopkinson D.,
 RA Edwards Y.,
 RT "Canine TCOF1: Cloning, Chromosome Assignment and Genetic analysis in
 RT dogs with different face and head types."
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ296287; CAC82741.1; -
 DR InterPro: IPR003993; treeacle.
 DR Pfam: PF03546; treeacle; 4.
 DR SEQUENCE 1422 AA; 145804 MW; B7381D3EDA79132 CRC64;

Query Match 6.6%; Score 151; DB 6; Length 1422;
 Best Local Similarity 21.5%; Pred. No. 0.41; Mismatches 171; Indels 118; Gaps 19;
 Matches 96; Conservative 61;

12 NTHNGLRMSIYKSTGNKGFEDGDAISKLDQEEYGALLIRNMQHIM-----GP 62
 109 STNSSVPGVLPSTSTKEGVAKTNKA-SKVNSTPHPAS--AKAVAHILSGRSPRKSAGP 165
 63 VTAIKLGEENKENTPRNVNLITPASMYAKAHKDKKKPORDSOAQSPYKSKSTHRIQ 122
 166 SANITLIVSETEEGS-----VPALGTTAKPGMASAN--QADSSSETS----- 205
 123 HNIDYLKHLSTKVKKIPSEEGSGYTDLOERGDNDISPESGDQPFKIDIPKGEATGPDL 182
 206 -----SSDETIVEYKASEKIYQAKAAGPYKGTGPGATPAP--- 243
 183 GKDIOTGFAGPSEASTHLDTKPKPYNEIPEREENGNTIGTRDETAKADAVSYLVEG 242
 244 -----PGKAGPSAAQ---KTEKP-----KEDS-----DSSEDSDEE----- 274
 243 SNDIMGSTNFKELPGREGNRVDAGSONAHQ-----GKVEFHYPPAPSKERKEG 291
 275 -EPPAKTLPQVKGSGKTPQVKAASAKESPRKGVPPVPGKV-----GPAAGAKKGAG 329
 292 SSDAESTNYNEIPKNGKS---TRKGVDSNRNOATLNEKORFPPSKGSQ-----GLPI 343
 330 EEDPDSTSEDSSEEAAPTAVPTSPYQAKPSGQNSQVRTASGPKVGPQAKGAPATPV 389
 344 PSRGIDNEIKNEDS-FNGSHENITTHGRKHYVPHRONNSTRNKMGPOGKSGWRGPH 402
 390 GKOEDESSSEESSEEGGAAPQAKSGK-----IPQVRAASGPAKGPQKAG-----PA 440
 403 SNRRFSRRDDSSSDSGSSSESD 428
 441 ATQAKAKMSKD-SSESSESESEEE 465

RESULT 14
 P70039 PRELIMINARY; PRT; 2829 AA.
 AC P70039; P79934;
 DT 01-FEB-1997 (TREMBlrel. 02, Created)
 DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE Adenomatous polyposis coli.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae;
 OC Xenopodinae; Xenopus.
 OC NCBI_TaxID=8353;
 RP SEQUENCE FROM N.A.
 RA Vlieminckx K., Wong E., Guger K., Gumbiner B.M.,
 RA Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U64442; ABA41671.1; -
 DR InterPro: IPR000225; Armdillo.
 DR Pfam: PF00514; Armdillo_seg; 6.
 DR SMART: SM00185; ARM; 5.
 DR SEQUENCE 2829 AA; 310878 MW; 8A2BABDB7706E496 CRC64;

Query Match 6.6%; Score 150; DB 13; Length 2829;
 Best Local Similarity 21.1%; Pred. No. 1.1; Mismatches 213; Indels 108; Gaps 22;
 Matches 104; Conservative 67;

29 KGEEDGDAI-SKLHQEEYGALLIR--NNQHIMGVYTAIKL-----LGEENKENT 77
 2101 KAIQEGANSIVSRH-QAAAGSLRSGSDSLSLKSGLSPFLTLDRKEKIT 2159
 78 PRNVNLITPASMYAKAHKDKKKPORDSOAQSPYKSKSTHRIQHNIDYLKHL--SKVK 135
 2160 SNKPKILKRAKSALENKKTTEEP-KGIGCKKVKYKSLITGKSRSSDSFSSHKQSOVOT 2218

RESULT 15	
09VAV4	
ID 09VAV4	PRELIMINARY;
AC 09VAV4; Q9VAV3;	PRT; 3060 AA.
DT 01-MAY-2000 (TREMBLrel. 13. Created)	
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)	
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)	
DE CG1340 protein.	
GN PPN OR CG1540 OR CG18436.	
OS Drosophila melanogaster (Fruit fly).	
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;	
OC pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;	
OC Ephydroidea; Drosophilidae; Drosophila.	
OX NCBI_TaxID=7227;	
NN [1]	

RC STRAIN-BERKELEY, ...
MEDLINE-20196006; PubMed-10731132;
RA Adams W.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Maniatis P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blaise J.R.G., Champe M., Pfeiffer B.D.,
RA Abril J.F., Abhyantari A., An H.-J., Andrews-Pfankoch C., Balgutin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhattacharjee D., Bolshakov S.,
RA Borkova D., Botchan M.R., Boulter J., Brockstein P., Brotier P.,
RA Burris K.C., Busam D.M.R., Butler H., Cadiheu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke S., Dahlke J., Davidson P., Davies P.,
RA de Fabios B., Delcher A., Deng Z., Mays A.D., Dew I., Dletz S.M.,
RA Dodson K., Doup L.E., Downes M., Dungan-Rocha S., Dunkov B.C., Dunn P.,
RA Durdin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Gang N.S., Gelbart W.M., Glasser K.,
RA Gloder A., Gong F., Gottell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaitani M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Kechum R.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lai Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mekulov G., Mlshina N.V., Moberly A., Morris J., Moshefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclob J.M.,

```

RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Relbert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamis I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spirdling A.C., Stapleton M., Strong R., Sun E.,
RA Svrtkals R., Tector C., Turner R., Ventier E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye Y., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster";
RL Science 287:2185-2195(2000).
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
DR EMBL: AE0033765; AAF56794.2; -.
DR EMBL: AE0033765; AAF56795.2; -.
DR HSSP: P12111; 2KNT.
DR FlyBase: FBgn0003137; Pm.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR003598; Ig_C2.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR002223; Kunitz_BptI.
DR InterPro: IPR000884; TSP1.
DR InterPro: IPR002221; WAP.
DR Pfam: PF00047; Ig_3.
DR Pfam: PF00014; Kunitz_BptI; 12.
DR Pfam: PF00090; tsp_1; 5.
DR Pfam: PF00095; wap; 1.
DR PRINTS: PR00759; BASICPTASE.
DR PRODOM: PD000222; Kunitz_BptI; 12.
DR SMART: SM00408; IGC2; 3.
DR SMART: SM00131; KU; 12.
DR SMART: SM00209; TSP1; 7.
DR SMART: SM00217; WAP; 1.
DR PROSITE: PS00317; 4_DISULFIDE_CORE; 1.
DR PROSITE: PS00280; BPTI_KUNITZ_1; 11.
DR PROSITE: PS00279; BPTI_KUNITZ_2; 12.
DR PROSITE: PS00022; EGF_1; UNKNOWN_1.
DR PROSITE: PS00092; TSP1; 3.
KW Alternative splicing; Immunoglobulin domain;
KW Serine protease inhibitor.
FT VARSPLIC 2803
FT FT 2803
FT VARSPLIC 2844
FT FT 2854
SQ VARSPLIC 2855 3060 MISSING (IN SHORT ISOFORM).
SQ SEQUENCE 3060 AA; 331579 MW; ACA31D3E855C7C0 CRC64;

Query Match 6.6%; Score 150; DB 5; Length 3060;
Best Local Similarity 19.1%; Pred. No. 1.2;
Matches 85; Conservative 58; Mismatches 173; Indels 126; Gaps 12

QY 7 ISNNENTHNLRLMSIYKSTGNGKGFEDGDAISKLDQETGALIRNNQHIMGPTAI 66
DB 644 LSEKONSEATEDEILPLRSTDKPIEDDE-----DCDEEDGIELISDLSDKESEYI 697
QY 67 KLLGEENEKENTPRNVLIITIPASMYAKAKSHDKKPPQDSQAQSPVSKSHRIRQHND 126
DB 698 DLEGTAKTETTP-----EAEEDLMQSSPPPIYDFBSTG----- 731
QY 127 YLKHLTKVKKIPSDPFGSGYITDLOEKQNDIDSPFGDGQFPKIDGKEATGPDLECKDI 186
DB 732 -----TEGSGY-DSESTTSGISIT-----EGSDDEFTSEA 762
QY 187 QTGFAGPSEASTLHDT-----KKPGYEIF-----EREENGNTIGTRDE-- 227
DB 763 SYDLSSSTDSGSTSDSISSDSSSISSDAISSEAPASSVSSDSSSTASTETTCVSDST 822
QY 228 ---TKKADAVDYSILVESNDIMGSTNFKELPGRGNRVADAGSONAHQKVEFIHPAPS 284
DB 823 DVSSTTEFASAEESTDVGASDSTGSTNASD-----ST 854
QY 285 KEKREGSSDAEESTNYEIRPKNGGSRKGVDSHNNQATLNKQRFPPKSGKSQGLPIP 344

```

Thu Apr 17 09:34:39 2003

us-09-700-696b-2.rspt

Page 9

```
Db      855 PESSTEASSSTDSTDSD---NSSNVSESTFEASSSVSDSND-----SDGSTGVSST 907
QY      345 SRCIDNEIKNEMDSFNGPSHENTITHGRKYHYVPHRQNNSTRNKGMPQGGSGWGRQPHSN 404
Db      908 TENSSDSTDATSDSTAASDSTDSTDQTTETTPESSTDSTESSTLDA-----SS 957
QY      405 RRFSSRRDDSSSSSDSGSSSED 428
Db      958 TTDASSTSESSSSTDGSSSTSN 981
```

Search completed: April 11, 2003, 12:03:23
Job time : 47 secs

Thu Apr 17 09:34:38 2003

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model
April 11, 2003, 12:01:02 : Search time 17 seconds
(Without alignments)
744.227 Million cell updates/sec

Title: US-09-700-696b-2
Perfect score: 2279
Sequence: 1 VNKEYSISNKENTHGLNLS.....RRDSSSSSSGSSSESDGD 430

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 262574 seqs, 29422922 residues 262574

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Issued Patents-AA:*
1: /cgn2.6/ptodata/1/1aa/5R-COMB.pep:*
2: /cgn2.6/ptodata/1/1aa/5R-COMB.pep:*
3: /cgn2.6/ptodata/1/1aa/5R-COMB.pep:*
4: /cgn2.6/ptodata/1/1aa/5R-COMB.pep:*
5: /cgn2.6/ptodata/1/1aa/5R-COMB.pep:*
6: /cgn2.6/ptodata/1/1aa/5R-COMB.pep:*

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	140	6.1	772	1 US-08-524-757-12	Sequence 12, Appl
2	136	6.0	1187	1 US-08-320-559-28	Sequence 28, Appl
3	136	6.0	1187	3 US-08-545-860D-28	Sequence 28, Appl
4	136	6.0	1187	5 PCT-US94-04496-28	Sequence 26, Appl
5	136	6.0	1210	1 US-08-320-559-26	Sequence 26, Appl
6	136	6.0	1210	1 US-08-545-860D-26	Sequence 26, Appl
7	136	6.0	1210	3 PCT-US94-04496-26	Sequence 11, Appl
8	135.5	5.9	723	1 US-07-814-964-11	Sequence 6, Appl
9	135.5	5.9	723	1 US-08-358-442-11	Sequence 11, Appl
10	135.5	5.9	723	1 US-08-328-809-6	Sequence 11, Appl
11	135.5	5.9	723	5 PCT-US93-11107-11	Sequence 2, Appl
12	134.5	5.9	1115	2 US-08-568-459A-2	Sequence 2, Appl
13	134.5	5.9	1115	2 US-08-487-826B-2	Sequence 2, Appl
14	134.5	5.9	1115	4 US-09-210-288-2	Patent No. 5198347
15	134.5	5.9	1115	6 5198347-6	Sequence 4, Appl
16	128	5.6	1235	1 US-08-910-925-4	Sequence 13, Appl
17	127.5	5.6	1235	1 US-08-118-101A-2	Sequence 16, Appl
18	126.5	5.6	1235	5 PCT-US93-07261-13	Sequence 4, Appl
19	126.5	5.6	1235	5 PCT-US93-07261-16	Sequence 2, Appl
20	125.5	5.5	1261	4 US-09-332-295-2	Sequence 2, Appl
21	125.5	5.5	1261	4 US-09-709-979-2	Sequence 3530, Ap
22	125.5	5.5	1183	4 US-09-134-001C-3530	Sequence 12, Appl
23	124.5	5.5	1183	4 US-08-999-774A-12	Sequence 5106, Ap
24	123	5.4	1177	4 US-08-134-001C-5106	Sequence 11, Appl
25	123	5.4	1588	4 PCT-US93-07261-11	Sequence 4568, Ap
26	123	5.4	1588	4 US-09-134-001C-4568	
27	122	5.4	690	4	

us-09-700-696b-2.ra1

28	121	5.3	1404	4 US-08-801-308-1	Sequence 1, Appl
29	120.5	5.3	1435	2 US-08-568-459A-4	Sequence 4, Appl
30	120.5	5.3	1435	2 US-08-487-826B-4	Sequence 4, Appl
31	120.5	5.3	1435	4 US-09-210-288-4	Sequence 9, Appl
32	120	5.3	1250	1 US-08-441-139-9	Sequence 7, Appl
33	120	5.3	2842	1 US-07-741-940-7	Sequence 7, Appl
34	120	5.3	2842	1 US-08-289-548A-7	Sequence 7, Appl
35	120	5.3	2842	1 US-08-452-654-7	Sequence 2, Appl
36	120	5.3	2843	4 US-08-449-731-7	Sequence 2, Appl
37	120	5.3	2843	1 US-07-741-940-2	Sequence 2, Appl
38	120	5.3	2843	1 US-08-289-548A-2	Sequence 2, Appl
39	120	5.3	2843	1 US-08-452-654-2	Sequence 2, Appl
40	120	5.3	2843	1 US-08-452-655B-7	Sequence 2, Appl
41	120	5.3	2843	1 US-08-370-235A-2	Sequence 2, Appl
42	120	5.3	2843	3 US-08-450-582-2	Sequence 7, Appl
43	120	5.3	2843	3 US-08-450-582-7	Sequence 7, Appl
44	120	5.3	2843	3 US-08-449-731-2	Sequence 2, Appl
45	120	5.3	2843	4	

ALIGNMENTS

RESULT 1
US-08-524-757-12 Application US/08524757
Sequence 12, Appl 5192634
Patent No. 5192634
GENERAL INFORMATION:
APPLICANT: CONAWAY, Ronald C.
APPLICANT: CONAWAY, Joan W.
APPLICANT: BRADSHAW, John W.
TITLE OF INVENTION: RNA Polymerase Transcription Factor
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESSES:
ADDRESSEE: RICHARDS, MEDLOCK & ANDREWS
STREET: 1201 Elm Street, Suite 4500
CITY: Dallas
STATE: TX
COUNTRY: US
ZIP: 75270-2197
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/524,757
CLASSIFICATION: 435
FILING DATE: PCT/US94/13621
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/13621
FILING DATE: 29-NOV-1994
PRIOR APPLICATION DATA: US 08/160087
APPLICATION NUMBER: 30-NOV-1993
FILING DATE: 30-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Harre, John A.
REGISTRATION/DOCKET NUMBER: 37,345
REFERENCE/DOCKET NUMBER: B35006CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (214) 939-4600
TELEFAX: (214) 939-4600
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 772 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-524-757-12
Query Match 6.1% Score 140: DB 1: Length 772:
Best Local Similarity 20.5% Pred No. 0.0011: Indels 144: Gaps 17:
Matches 87: Conservative

OY 23 PKSTGNKGFEDGDDAISKLDHDEEYGAALIRNMHONGPVTALIKLGEENKENTPRNVL
 Db 220 PKGHSNAPFODRIGASOEHLGEPHKGIVSONKEH
 OY 83 NITASNNYAKAHSKDKKPPORDSOAKSPYKSKSTRIOHNDYILKHSKPKIPSPDE 245
 Db 246 KSHMDKRVADKSDENASVYSREKSH
 OY 143 GSGCYDLOERGNDISPSSGGOPEFKDIPGKGATPDLEGDIOTGFAGPSEAS-THL 201
 Db 278 FENRRPPSGNAEKRP-PSSGKKEKEKRESSSLKKCLPPEASADNHL 325
 OY 202 DTRKRGVNEIPEREENGNGTIGTRDETAADAVVSLVEGSNDIM-GSTNFKEL 255
 Db 326 KRPKRD-PERAK-LKSKOGIDSEFDYG-KGAGDLPKYKEKGSNNLTKP 372
 OY 256 PGRGNRVADSONAHQGV-EFHP-PAPSKERKESGSDAAS 298
 Db 373 EGKVTNLDKRSIGS-LKRVETDMEDEFEQPTWSPESYSTLDOPKRRKKIVKT-SATA 430
 OY 299 TYNNEIPKNGKSTKRGVDSNRQATLNEKORFSPKSGQL-ITHGKRYTP- 378
 Db 431 LGDKGLKNDKSTGKNDVOKLPRKNTKSEKPAADLAKRKVPDVLPLDPLPA 490
 OY 342 PIRSGDNEITKNEKMDSPNGSHEN-ITHGKRYTP- 378
 Db 491 IQANRPLPSELISSFOPKRAFSSPOEEBAGFTGRNMSKMQVSSSKCAVLPKMT 550
 OY 379 -HRO 381
 Db 551 LHQO 554

RESULT 2
 US-08-320-559-28
 Sequence 28, Application US/08320559
 Patent No. 5633135
 GENERAL INFORMATION:
 APPLICANT: Croce, Carlo
 TITLE OF INVENTION: Diagnostics, Therapeutics and Methods for
 TITLE OF INVENTION: Detection and Treatment of Acute Leukemias
 NUMBER OF SEQUENCES: All-1 Region
 CORRESPONDENCE ADDRESS:
 ADDRESS: Woodcock Washburn Kurtz Mackiewicz & No. 5633135
 CITY: Philadelphia
 STATE: PA
 COUNTRY: USA
 ZIP: 19103
 COMPUTER READABLE FORM:
 MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
 COMPUTER: IBM PS/2
 OPERATING SYSTEM: PC-DOS
 SOFTWARE: WORDPERFECT 5.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/320,559
 FILING DATE: US/08/320,559
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/08/062,443
 FILING DATE: 14 MAY 1993
 APPLICATION NUMBER: US/07/971,094
 FILING DATE: 30-OCT-92
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/07/888,830
 FILING DATE: 27-MAY-92
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/07/805,093

FILING DATE: 11-DEC-91
 ATTORNEY/AGENT INFORMATION:
 NAME: DeLuca, Mark
 REGISTRATION NUMBER: 33,229
 REFERENCE/DOCKET NUMBER: TJU-0855
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (215) 568-3439
 TELEFAX: (215) 568-3100
 INFORMATION FOR SEQ ID NO: 28:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1187 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-320-559-28

Query Match Best Local Similarity 6.0%, Score 136, DB 1, Length 1187,
 Matches 86; Conservative 52; Mismatches 166; Indels 114; Gaps 14,
 OY 92 AKASDKKKPOROS-ONASPYKSKSTRIOHNDYILKHSKYK 135
 Db 530 SQHSESDPPPKSSKAPAPPEAPHPKRCSCQSPAQOPPOROTVGTGKPK 584
 OY 136 KIPDFESGYTDIOERGNDISPSSGGOPEFKDIP-GKGE-ATGPD 181
 Db 585 PKNASARAGSRTSLOGBERGLPYGSRDQTSKDKPKVKTGKPRRAASNEKRPVAVPSS 644
 OY 182 EGKDIOGFAGPSEAS-THLDTKRPY-NEIPEREENGNGTIGTRDE 227
 Db 645 EKKKHKSSLPAPSKALSGEPKAKONVEDRTPEHVALPLTESGPHSGSRTSGGQA 704
 OY 228 TAKEADAVVSLVEGSNDIMSTNFKELPEREENK-DAGSONAHQGV 276
 Db 705 VVVOEDSRKRLPLRLDTYKLSLPLDTPPOSILKTLTLRLIPORPKSRRKAE 764
 OY 277 FHYPPAPSKERKRGSSDAESTYNNEIPKNGKSTKRGVDSNRQATLNEKORFSPK 336
 Db 765 DKOPKAKKSSSRSSSS-SKLAKRKGEARDCD-MKKIR- 805
 OY 337 KSQGLPIPSGDLNETKNEKMDSPNGSHENIT-HOKHYVPHRONNSTRNK 388
 Db 806 LEKEIKSSSSSSSHKESSTKTPKSPSSQSKKEMLPVAVSSSQK 854
 OY 389 GMPQKGS-WGROP-HSNRPSSRRRDDSSSDSGSSSE-SOGD 430
 Db 855 AKPALKRSRRADTCODPPKASASTKSNHKDSSIPKORVVEGKSSSSSHKSSGD 912

RESULT 3
 US-08-545-860D-28
 Sequence 28, Application US/08545860D
 Patent No. 6040140
 GENERAL INFORMATION:
 APPLICANT: Croce, Carlo
 TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
 TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
 NUMBER OF SEQUENCES: 94
 CORRESPONDENCE ADDRESS:
 ADDRESS: Woodcock, Washburn, Kurtz, Mackiewicz &
 STREET: One Liberty Place, 46th floor
 CITY: Philadelphia
 STATE: Pennsylvania
 COUNTRY: USA
 ZIP: 19103
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentln Release #1 0, Version #1 25

;;
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/545,860D
FILING DATE: 07-MAR-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/04496
FILING DATE: 22-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/10930
FILING DATE: 09-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/327,392
FILING DATE: 19-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/320,559
FILING DATE: 11-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/062,443
FILING DATE: 14-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/971,094
FILING DATE: 30-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/888,839
FILING DATE: 27-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/805,093
FILING DATE: 11-DEC-1991
ATTORNEY/AGENT INFORMATION:
NAME: Deluca Esq., Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: TJU-1262
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 1187 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-545-860D-28

Query Match 6.0%; Score 136; DB 3; Length 1187;
Best Local Similarity 20.6%; Pred. No. 0.0045;
Matches 86; Conservative 52; Mismatches 166; Indels 114; Gaps 14;
QY 92 AKHSDKKKPPORDS-----QAQSPVKSSTHRIQHNIDYKHLKSKYK 135
DB 530 SQHSSSKPPPKSSSKAPRAPPEARHPGKRSQKSPAOOEPPROTGTGKPK-----K 584
QY 136 KISDDEGSGYTDLOERGNDISPFSGDGPFPKIDP---GKGE-----ATGPD 181
DB 585 PYVASRAGSRTSLQGERBGLPYGSRDQTSKDKPKVTKGPRRAASNEKPPAVPSS 644
QY 182 EKGDIOTGAGPSEAS-----THLDTKKPGY-----NEIPREENGNTTIGTRDE 227
DB 645 EKKHKSLSLPAPSKALSGEPKADVEDRTPEHFALVPLTESGPHSGSGTSQCRQA 704
QY 228 TAKEADAVNVSLEGSNDIMGSTNFKELPGREGNRY-----DAGSONAHGKYE 276
DB 705 VVQEDSRKDRPLPLRDTKLLSPLRDTPPPSLWKITLDSLRIPOPFGKGSRRKAE 764
QY 277 FHYPPAPSKERKREGSSDAESTNYNEIPKNGSGTRKGVDSNRQATLNEKORPPSG 336
DB 765 DKOPRAGKHSSEKSSDS-----SKLAKKRGAEARCD-----NKKIR----- 805
QY 337 KSGGLPIPRGLDNETKNEMDSFNKPSHENIIT-----HGKHYVYPHRONNSTRNK 388
DB 806 -----LEKEIKSSSSSSSHKSSSKTKPSPSSQSKKMLPPPPVSSSSQKP 854
QY 389 GMPQKGS-----WGROP-----HSNRFPSSRRRDSSESDDGSSSE---SNGD 430

DB 855 AKPALKRSREADTCGDDPPKASSTKSNHKDSDIPKQREVEGKSGSSSEHKSSGD 912
RESULT 4
PCT-US94-04496-28
Sequence 28, Application PC/TUS9404496
GENERAL INFORMATION:
APPLICANT: Croce, Carlo
TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
NUMBER OF SEQUENCES: 86
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
STREET: One Liberty Place, 46th floor
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/04496
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Deluca Esq., Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: TJU-1242
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 1187 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-04496-28

Query Match 6.0%; Score 136; DB 5; Length 1187;
Best Local Similarity 20.6%; Pred. No. 0.0045;
Matches 86; Conservative 52; Mismatches 166; Indels 114; Gaps 14;
QY 92 AKHSDKKKPPORDS-----QAQSPVKSSTHRIQHNIDYKHLKSKYK 135
DB 530 SQHSSSKPPPKSSSKAPRAPPEARHPGKRSQKSPAOOEPPROTGTGKPK-----K 584
QY 136 KISDDEGSGYTDLOERGNDISPFSGDGPFPKIDP---GKGE-----ATGPD 181
DB 585 PYVASRAGSRTSLQGERBGLPYGSRDQTSKDKPKVTKGPRRAASNEKPPAVPSS 644
QY 182 EKGDIOTGAGPSEAS-----THLDTKKPGY-----NEIPREENGNTTIGTRDE 227
DB 645 EKKHKSLSLPAPSKALSGEPKADVEDRTPEHFALVPLTESGPHSGSGTSQCRQA 704
QY 228 TAKEADAVNVSLEGSNDIMGSTNFKELPGREGNRY-----DAGSONAHGKYE 276
DB 705 VVQEDSRKDRPLPLRDTKLLSPLRDTPPPSLWKITLDSLRIPOPFGKGSRRKAE 764
QY 277 FHYPPAPSKERKREGSSDAESTNYNEIPKNGSGTRKGVDSNRQATLNEKORPPSG 336
DB 765 DKOPRAGKHSSEKSSDS-----SKLAKKRGAEARCD-----NKKIR----- 805
QY 337 KSGGLPIPRGLDNETKNEMDSFNKPSHENIIT-----HGKHYVYPHRONNSTRNK 388
DB 806 -----LEKEIKSSSSSSSHKSSSKTKPSPSSQSKKMLPPPPVSSSSQKP 854

QY 389 GMPGKGS-----WGQP-----HSNRRFSRRRDDSSSSDSSGSSSE---SDGD 430
Db 855 AKPALKRSRRADTCGDDPPKASSTKSNHKDSSIPQRVYEGKSGSSSEHKSSGD 912

RESULT 5

US-08-320-559-26
Sequence 26, Application US/08320559
Patent No. 563135
GENERAL INFORMATION:
APPLICANT: Croce, Carlo
TITLE OF INVENTION: Diagnostics, Therapeutics and Methods for
TITLE OF INVENTION: Detection and Treatment of Acute Leukemias
TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 563135rls
STREET: One Liberty Place - 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/320,559
FILING DATE: 11-DEC-91
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/062,443
FILING DATE: 14 MAY 1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/971,094
FILING DATE: 30-OCT-92
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/888,830
FILING DATE: 27-MAY-92
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/805,093
FILING DATE: 11-DEC-91
ATTORNEY/AGENT INFORMATION:
NAME: Deluca, Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: TJU-0855
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 1210 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-320-559-26

Query Match 6.0%; Score 136; DB 1; Length 1210;
Best Local Similarity 20.6%; Pred. No. 0.0046;
Matches 86; Conservative 52; Mismatches 166; Indels 114; Gaps 14;

QY 92 AKASHKDKKPKQDS-----QAQKSPYKSKSTHRIQHNIDYLKHLKRYK 135
Db 553 SQHSESKDPKPPKSSAPRAPAPHPGKRSCKSPAQCPPOQRTVGTQPK-----K 607
QY 136 KIPDDESSGYTDIOERDNDISPESGSGGPFKIP---GKGE-----ATGPD 181
Db 608 PVKASARAGSTSLQGEREPGLPPGSRDQTSKDKPVKTKGRPRAAASNPKPAVPPSS 667
QY 182 EGKDIOTGTGAPSEAES-----THLDTKKPGY-----NEIPEREENGNTIGTRDE 227

Db 668 EKKHKSLAPSKALSGPEPAKQNVEDRTPEHVALVLTESQGPSPHSGSGSRTSGCROA 727
QY 228 TAKEADAVDSLYEGSNDINGSTNFKELPGREGNRV-----DASQNAHQKVE 276
Db 728 VVVQEDSRKRDPLPLDRTLLSPRLDTPPPQSLMVKITLDSLRIPOPPGKSGRQRAE 787
QY 277 FHVPAPSKKRRKSGSDAESTNVEIPKNGKSTKGVDSNRNQTINERKPFSGK 336
Db 788 DKPPAKKHSSEKSSDSS-----SKLAKKRGAEARCD-----NKTR----- 828
QY 337 KSOGLPIPSKGLDNEIKNEMDSFNGPSHENIT-----HGKRYHVPDRONSTRNK 388
Db 829 -----LEKETISQSSSSSSSHKESKTKPSRSPSSOSKEMLPPIVSSSSQXP 877
QY 389 GMPGKGS-----WGQP-----HSNRRFSRRRDDSSSSDSSGSSSE---SDGD 430
Db 878 AKPALKRSRRADTCGDDPPKASSTKSNHKDSSIPQRVYEGKSGSSSEHKSSGD 935

RESULT 6

US-08-545-860D-26
Sequence 26, Application US/08545860D
Patent No. 6040140
GENERAL INFORMATION:
APPLICANT: Croce, Carlo
TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1 Region
NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn, Kurtz, Mackiewicz &
STREET: One Liberty Place, 46th floor
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/545,860D
FILING DATE: 07-MAR-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/04496
FILING DATE: 22-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/10930
FILING DATE: 09-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/327,392
FILING DATE: 19-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/320,559
FILING DATE: 11-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/062,443
FILING DATE: 14-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/971,094
FILING DATE: 30-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/888,839
FILING DATE: 27-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/805,093
FILING DATE: 11-DEC-1991
ATTORNEY/AGENT INFORMATION:

NAME: Deluca Esq., Mark
 REGISTRATION NUMBER: 33,229
 REFERENCE/DOCKET NUMBER: TJU-1262
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (215) 568-3100
 TELEFAX: (215) 568-3439
 INFORMATION FOR SEQ ID NO: 26:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1210 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-545-860b-26

Query Match 6.0%; Score 136; DB 3; Length 1210;
 Best Local Similarity 20.6%; Pred. No. 0.0046;
 Matches 86; Conservative 52; Mismatches 166; Indels 114; Gaps 14;

92 AKASRKKKKRPPORDS-----QAQKSPVSKSTHRIQHNIDYKHLKSKVK 135
 553 SQEHSSEKDPPEKSSSKAPAPPEAPHPGKRCQKSPAQEPPOQVGTQPK-----K 607
 136 KIPDFEGSGYTDLQERGNDISPESGQGPFPKIP---GKGE-----ATGPD 181
 608 PVKASARAGSRISLQGEREPGLPYGSRDQTSKDKPKVTKGPRRAASNEPKPAVPSS 667
 182 EKGDIQTGFAGPSEAES-----THLDTKKPGY-----NEIPEREENGNTIGTDE 227
 668 EKKKHKSSLPAPSKALSGPEPAKDNVEDRTPEHFALVPLTESQGPSPHSGSRTSCQRA 727
 228 TAKEADAVDSLVESGNDIMSGTNFKELPREGNRY-----DAGSONAHQKVE 276
 728 VVQEDSRKRRLPLRLDTYKLLSLRDTPPQSLMWKITLDSLRIPOPPGKSGRQKAE 787
 277 FHYPAPEKREKREGSSDAESTVNEIPKNGKSTRKGVDSNRNQTALNEKORFPKSG 336
 788 DKQPPAGKKHSEKSSSS-----SKLAKRKGAEARDCD-----NKKIR----- 828
 337 KSGGLPIPSGLDNEIKEMDSFNGPSHENIT-----HGKRYHYVPHRONNSTRNK 388
 829 -----LEKEIKSSSSSSSHKESKTKPSPSSQSKKMLPPPVSSSQKP 877
 389 GMPQKGS-----WGRP-----HSNRRFSSRRDDSSSSDSGSSSE---SDGD 430
 878 AKPALKRSRREADTCGDDPPKASSTKSNHKDSSIPKQRRVEGKGRSSSEHKGSSGD 935

RESULT 7
 PCT-US94-04496-26
 Sequence 26, Application PC/TUS9404496
 GENERAL INFORMATION:
 APPLICANT: Croce, Carlo
 APPLICANT: Canaan, Eli
 TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
 TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
 NUMBER OF SEQUENCES: Resulting from Chromosome Abnormalities in the All-1
 CORRESPONDENCE ADDRESS: 86
 ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
 ADDRESS: Norris
 STREET: One Liberty Place, 46th floor
 CITY: Philadelphia
 STATE: Pennsylvania
 COUNTRY: USA
 ZIP: 19103
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US94/04496
 FILING DATE:

CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Deluca Esq., Mark
 REGISTRATION NUMBER: 33,229
 REFERENCE/DOCKET NUMBER: TJU-1242
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (215) 568-3100
 TELEFAX: (215) 568-3439
 INFORMATION FOR SEQ ID NO: 26:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1210 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 PCT-US94-04496-26

Query Match 6.0%; Score 136; DB 5; Length 1210;
 Best Local Similarity 20.6%; Pred. No. 0.0046;
 Matches 86; Conservative 52; Mismatches 166; Indels 114; Gaps 14;

92 AKASRKKKKRPPORDS-----QAQKSPVSKSTHRIQHNIDYKHLKSKVK 135
 553 SQEHSSEKDPPEKSSSKAPAPPEAPHPGKRCQKSPAQEPPOQVGTQPK-----K 607
 136 KIPDFEGSGYTDLQERGNDISPESGQGPFPKIP---GKGE-----ATGPD 181
 608 PVKASARAGSRISLQGEREPGLPYGSRDQTSKDKPKVTKGPRRAASNEPKPAVPSS 667
 182 EKGDIQTGFAGPSEAES-----THLDTKKPGY-----NEIPEREENGNTIGTDE 227
 668 EKKKHKSSLPAPSKALSGPEPAKDNVEDRTPEHFALVPLTESQGPSPHSGSRTSCQRA 727
 228 TAKEADAVDSLVESGNDIMSGTNFKELPREGNRY-----DAGSONAHQKVE 276
 728 VVQEDSRKRRLPLRLDTYKLLSLRDTPPQSLMWKITLDSLRIPOPPGKSGRQKAE 787
 277 FHYPAPEKREKREGSSDAESTVNEIPKNGKSTRKGVDSNRNQTALNEKORFPKSG 336
 788 DKQPPAGKKHSEKSSSS-----SKLAKRKGAEARDCD-----NKKIR----- 828
 337 KSGGLPIPSGLDNEIKEMDSFNGPSHENIT-----HGKRYHYVPHRONNSTRNK 388
 829 -----LEKEIKSSSSSSSHKESKTKPSPSSQSKKMLPPPVSSSQKP 877
 389 GMPQKGS-----WGRP-----HSNRRFSSRRDDSSSSDSGSSSE---SDGD 430
 878 AKPALKRSRREADTCGDDPPKASSTKSNHKDSSIPKQRRVEGKGRSSSEHKGSSGD 935

RESULT 8
 US-07-814-964-11
 Sequence 11, Application US/07814964
 Patent No. 5359047
 GENERAL INFORMATION:
 APPLICANT: Donahue, Brian A.
 APPLICANT: Toney, Jeffrey H.
 APPLICANT: Bruhn, Suzanne L.
 APPLICANT: Pil, Pieter M.
 APPLICANT: Brown, Steven
 APPLICANT: Kelleit, Patti
 APPLICANT: Essigmann, John M.
 TITLE OF INVENTION: DNA Structure Specific Recognition
 TITLE OF INVENTION: Protein and Uses Therefor
 NUMBER OF SEQUENCES: 13
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
 STREET: 2 Militia Drive
 CITY: Lexington
 STATE: MA
 COUNTRY: USA
 ZIP: 02173
 COMPUTER READABLE FORM:

[illegible]

```

;
; OTHER INFORMATION: /label= HMG-bor
;
; FEATURE:

```


[illegible]

```

RESULT 12
US-08-568-459A-2
Sequence 2, Application US/08568459A
Patent No. 5849306
GENERAL INFORMATION:
APPLICANT: Sim, Kim I.
APPLICANT: Chltnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEIN
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/568,459A
FILING DATE: 07-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelisen, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CPI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1115 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium vivax
US-08-568-459A-2
Query Match 5.9%; Score 134.5; DB 2; Length 1115;
Best Local Similarity 20.0%; Pred. No. 0.0055;
Matches 97; Conservative 80; Mismatches 220; Indels 89; Gaps 21;

```

```

QY 2 NKEYSINKENTHNGLSIYKSTGKGFEDGDAISKLDQDEYGAALIRNNQHTMG 61
DB 500 NKFIYVKAKEKQVAGIYTPY-----DILKQ--ELDEFNEVARENEIKNRDG 544
QY 62 PVTAIKLIG-EBNKENTPRNVNLIIPASMYAKASHKDKKRPORDSOAKSPVSKSTHR 120
DB 545 AYIELCVCSVEEAKKNTQEVYTNVDNAKSAQ-TNSNPIISQPVDSKAKEYP--GDSTH- 600
QY 121 IOHNDILKHLISKV-KKIPSPFEGSGYTDLO--ERGDNDISPFSGDQPKPIPGKGE- 175
DB 601 --GAVNSGQDSSTTGKAVTGDQNGNQTPAESDVQYRSDIAESVAKNVDPQKSVSKRSD 658
QY 176 --ATGPDLGKDIQTGFAGPSEAEETHLDTKKRG-----YNEI 211
DB 659 TASVTGIAEAGKE-NLGASNSRPEST-VEANSFGDDTVNSASIPVYSGENPLVTPYNGL 716
QY 212 PEREENGK-----TIGTRDETAKEADAVDVSLEGSNDIMGSTNFKELPREGN 261
DB 717 RHSKDNDSDGPAESMANPDNSKGETGKGQDNDMAKATKDDSSNSDGT-----SATGD 771
QY 262 RVDAGSONAHQKVEFHPAPAPSKERKREGSS--DAESTNYNEIIPKNGKST---RKG 315
DB 772 TTDAVDREINKGVPEDRDKTVSGKDGGEEDNSANKDAATVYGEDRIENSGAGSTNDRSK 831
QY 316 VDHNRNQAATLNEKOR--FPSKSGSGLPIPSRG--LDNEIKNEMDSFNGSHENITTHG 371
DB 832 NDEKNGASTPDSKQSEDAATLSTKTESLESTESGDRTTNDTNTNLEKNGKKEKLDQKHD 891
QY 372 RKHYVPHRONNSTR-----NKGMPQKGSWGRQPHSNR-RFSRRRDDSSSSDSC 422
DB 892 FKNDTPNEEPNSDQTTDAEGHDSIKNDKAE--RRKHMKDTFTKNTNSHHLNSNNL 949
QY 423 SSESSED 428
DB 950 SNGKLD 955

RESULT 13
US-08-487-826B-2
; Sequence 2, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; TITLE OF INVENTION: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CPI
; TELECOMMUNICATION INFORMATION:

```

```

; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1115 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium vivax
; US-08-487-826B-2

Query Match 5.9%; Score 134.5; DB 2; Length 1115;
Best Local Similarity 20.0%; Pred. No. 0.0055;
Matches 97; Conservative 80; Mismatches 220; Indels 89; Gaps 21.

QY 2 NKEYSINKENTHNGLSIYKSTGKGFEDGDAISKLDQDEYGAALIRNNQHTMG 61
DB 500 NKFIYVKAKEKQVAGIYTPY-----DILKQ--ELDEFNEVARENEIKNRDG 544
QY 62 PVTAIKLIG-EBNKENTPRNVNLIIPASMYAKASHKDKKRPORDSOAKSPVSKSTHR 120
DB 545 AYIELCVCSVEEAKKNTQEVYTNVDNAKSAQ-TNSNPIISQPVDSKAKEYP--GDSTH- 600
QY 121 IOHNDILKHLISKV-KKIPSPFEGSGYTDLO--ERGDNDISPFSGDQPKPIPGKGE- 175
DB 601 --GAVNSGQDSSTTGKAVTGDQNGNQTPAESDVQYRSDIAESVAKNVDPQKSVSKRSD 658
QY 176 --ATGPDLGKDIQTGFAGPSEAEETHLDTKKRG-----YNEI 211
DB 659 TASVTGIAEAGKE-NLGASNSRPEST-VEANSFGDDTVNSASIPVYSGENPLVTPYNGL 716
QY 212 PEREENGK-----TIGTRDETAKEADAVDVSLEGSNDIMGSTNFKELPREGN 261
DB 717 RHSKDNDSDGPAESMANPDNSKGETGKGQDNDMAKATKDDSSNSDGT-----SATGD 771
QY 262 RVDAGSONAHQKVEFHPAPAPSKERKREGSS--DAESTNYNEIIPKNGKST---RKG 315
DB 772 TTDAVDREINKGVPEDRDKTVSGKDGGEEDNSANKDAATVYGEDRIENSGAGSTNDRSK 831
QY 316 VDHNRNQAATLNEKOR--FPSKSGSGLPIPSRG--LDNEIKNEMDSFNGSHENITTHG 371
DB 832 NDEKNGASTPDSKQSEDAATLSTKTESLESTESGDRTTNDTNTNLEKNGKKEKLDQKHD 891
QY 372 RKHYVPHRONNSTR-----NKGMPQKGSWGRQPHSNR-RFSRRRDDSSSSDSC 422
DB 892 FKNDTPNEEPNSDQTTDAEGHDSIKNDKAE--RRKHMKDTFTKNTNSHHLNSNNL 949
QY 423 SSESSED 428
DB 950 SNGKLD 955

RESULT 14.
US-09-210-288-2
; Sequence 2, Application US/09210288
; Patent No. 6392026
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; TITLE OF INVENTION: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th floor
; CITY: Newport Beach

```


STATE: California
 COUNTRY: US
 ZIP: 92660
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/210,288
 FILING DATE:
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Fuller, Michael
 REGISTRATION NUMBER: 36,516
 REFERENCE/DOCKET NUMBER: NIH121.1FMDV1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 235-8550
 TELEFAX: (619) 235-0176
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1115 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHEICAL: NO
 ORIGINAL SOURCE:
 ORGANISM: Plasmodium vivax
 US-09-210-288-2

Query Match 5.9%; Score 134.5; DB 4; Length 1115;
 Best Local Similarity 20.0%; Pred. No. 0.0055;
 Matches 97; Conservative 80; Mismatches 220; Indels 89; Gaps 21;

QY 2 NKEYSISNRENTNGLRMSIYKSTGNKGFEDGDDAISKLHDOEYGAALLINNMQHIMG 61
 DB 500 NKFIYVNAEKVQTAGIVTPY-----DILKO--ELDEFNEVAFENEINKRDG 544
 QY 62 PVTAIKLLG--EENKENTPRNVLTIPASMYAKAHSKDKKPPORDSOAKSPYKSKSTR 120
 DB 545 AYIELCVSYVEAKKNTQEVYTWDNAAKSOA--TNSNPISQPVDSKAEKVP--GDSTH- 600
 QY 121 IQHNIDYLKHLISKY-KKIPDFEGSGYTDLO---ERGDNDISPFSGDGPFPKDIIPKGE- 175
 DB 601 --GNVNSGQDSSTTGKAVTGGQNGNQTPAESDVQSRDLASVSAAKNVDPQKSVKRSDD 658
 QY 176 --ATGPDLEKDIOTGFAGPSEAEETHLDTKPKG-----YNEI 211
 DB 659 TASVTGIAEAGKE-NLGASNSRPEST--VEANSPGDDTVNSASIPVSGENPLVTPYNGL 716
 QY 212 PERENGON-----TIGTRDETAKEDADVDSILVEGSNDIMGSTNFKELPGREGN 261
 DB 717 RHKSNDSDSGPAESMANPDSNSKGETGKGQNDMAKATKSSNSDGT-----SATGD 771
 QY 262 RVDAGSONAHOGKVEFHYPPAPSKERKEGSS---DAESTNYNEIIPKNGKST---RKG 315
 DB 772 TTDAVDREINKGVPEDRDKTVGSKDGGEDNSANKDAATVVGEDRIRENSAGSSTNDRSK 831
 QY 316 VDHNNRNOATLNEKOR--FPSKGSQGLPIPSRG--LDNEIKNEMDSFNGPSHENITTHG 371
 DB 832 NDEKNGASTPDSKQSEDATAALSKTESTESGDRTTNTDNTNLEKNGKGEKDIQKHD 891
 QY 372 RRYHYVPHRONNSTR-----NKGMPQKGSWGRQPHSNR--RFSRRRDDSSSSDSG 422
 DB 892 FKSNDTPNEEPNSDQTTDAEGHSDRSIKNDKAE--RRKHMKDTFTKNTNSHLLNSNNL 949
 QY 423 SSSESD 428
 DB 950 SNGKLD 955

RESULT 15

5198347-6
 Patent No. 5198347
 APPLICANT: Miller, LOUIS H.; ADAMS, JOHN H.; KASLOW,
 DAVIS C.; FANG, XIANGDOUG
 TITLE OF INVENTION: DNA ENCODING PLASMOIDIUM VIVAX AND
 PLASMODIUM KNOWLES DUFFY RECEPTOR
 NUMBER OF SEQUENCES: 27
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/554,837
 FILING DATE: 20-JUL-1990
 SEQ ID NO: 6:
 LENGTH: 1115
 5198347-6

Query Match 5.9%; Score 134.5; DB 6; Length 1115;
 Best Local Similarity 20.0%; Pred. No. 0.0055;
 Matches 97; Conservative 80; Mismatches 220; Indels 89; Gaps 21;

QY 2 NKEYSISNRENTNGLRMSIYKSTGNKGFEDGDDAISKLHDOEYGAALLINNMQHIMG 61
 DB 500 NKFIYVNAEKVQTAGIVTPY-----DILKO--ELDEFNEVAFENEINKRDG 544
 QY 62 PVTAIKLLG--EENKENTPRNVLTIPASMYAKAHSKDKKPPORDSOAKSPYKSKSTR 120
 DB 545 AYIELCVSYVEAKKNTQEVYTWDNAAKSOA--TNSNPISQPVDSKAEKVP--GDSTH- 600
 QY 121 IQHNIDYLKHLISKY-KKIPDFEGSGYTDLO---ERGDNDISPFSGDGPFPKDIIPKGE- 175
 DB 601 --GNVNSGQDSSTTGKAVTGGQNGNQTPAESDVQSRDLASVSAAKNVDPQKSVKRSDD 658
 QY 176 --ATGPDLEKDIOTGFAGPSEAEETHLDTKPKG-----YNEI 211
 DB 659 TASVTGIAEAGKE-NLGASNSRPEST--VEANSPGDDTVNSASIPVSGENPLVTPYNGL 716
 QY 212 PERENGON-----TIGTRDETAKEDADVDSILVEGSNDIMGSTNFKELPGREGN 261
 DB 717 RHKSNDSDSGPAESMANPDSNSKGETGKGQNDMAKATKSSNSDGT-----SATGD 771
 QY 262 RVDAGSONAHOGKVEFHYPPAPSKERKEGSS---DAESTNYNEIIPKNGKST---RKG 315
 DB 772 TTDAVDREINKGVPEDRDKTVGSKDGGEDNSANKDAATVVGEDRIRENSAGSSTNDRSK 831
 QY 316 VDHNNRNOATLNEKOR--FPSKGSQGLPIPSRG--LDNEIKNEMDSFNGPSHENITTHG 371
 DB 832 NDEKNGASTPDSKQSEDATAALSKTESTESGDRTTNTDNTNLEKNGKGEKDIQKHD 891
 QY 372 RRYHYVPHRONNSTR-----NKGMPQKGSWGRQPHSNR--RFSRRRDDSSSSDSG 422
 DB 892 FKSNDTPNEEPNSDQTTDAEGHSDRSIKNDKAE--RRKHMKDTFTKNTNSHLLNSNNL 949
 QY 423 SSSESD 428
 DB 950 SNGKLD 955

Search completed: April 11, 2003, 12:04:10
 Job time : 24 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2003, 12:02:21 ; Search time 19 Seconds
(without alignments)
1383.605 Million cell updates/sec

Title: US-09-700-696b-2
Perfect score: 2279
Sequence: 1 VNKEYSISNKENHNGLRMS.....RRDSESSSDSGSSSESDGD 430

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 248812 seqs, 61136040 residues

Total number of hits satisfying chosen parameters: 248812

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA.*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PC7_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/PC7US_PUBCOMB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2269	99.6	525	10	US-09-814-550-2
2	514	22.6	97	9	US-09-812-485A-1
3	246	10.8	47	9	US-09-812-485A-2
4	235.5	10.3	47	9	US-09-812-485A-4
5	229	10.0	47	9	US-09-812-485A-3
6	218	9.6	40	9	US-09-812-485A-23
7	216	9.5	44	9	US-09-812-485A-6
8	215	9.4	44	9	US-09-812-485A-5
9	214.5	9.4	45	9	US-09-812-485A-27
10	198.5	8.7	41	9	US-09-812-485A-8
11	190.5	8.4	40	9	US-09-812-485A-24
12	188.5	8.3	40	9	US-09-812-485A-32
13	182	8.0	37	9	US-09-812-485A-7
14	180	7.9	38	9	US-09-812-485A-10
15	163.5	7.2	35	9	US-09-812-485A-25
16	157.5	6.9	35	9	US-09-812-485A-28
17	153.5	6.7	33	9	US-09-812-485A-30
18	150	6.6	32	9	US-09-812-485A-12
19	148	6.5	32	9	US-09-812-485A-17

20	148	6.5	1907	10	US-09-785-770A-16	Sequence 16, Appl
21	146.5	6.4	33	9	US-09-812-485A-31	Sequence 31, Appl
22	144.5	6.3	665	9	US-09-820-843A-107	Sequence 107, App
23	142.5	6.3	31	9	US-09-812-485A-33	Sequence 33, Appl
24	142.5	6.3	31	9	US-09-812-485A-36	Sequence 36, Appl
25	137.5	6.0	30	9	US-09-812-485A-26	Sequence 26, Appl
26	134.5	5.9	1115	9	US-10-153-273-2	Sequence 29, Appl
27	129.5	5.7	28	9	US-09-812-485A-29	Sequence 14, Appl
28	129	5.7	28	9	US-09-812-485A-14	Sequence 64, Appl
29	127	5.6	322	9	US-09-746-660A-64	Sequence 5661, Ap
30	127	5.6	718	9	US-09-738-628-5661	Sequence 7, Appl
31	127	5.6	1400	10	US-09-764-176-7	Sequence 49, Appl
32	126	5.5	23	9	US-09-812-485A-49	Sequence 2, Appl
33	125.5	5.5	1143	10	US-09-924-154-14	Sequence 37, Appl
34	125.5	5.5	1261	12	US-10-147-268-2	Sequence 35, Appl
35	124.5	5.5	27	9	US-09-812-485A-35	Sequence 9, Appl
36	124	5.4	28	9	US-09-812-485A-9	Sequence 3, Appl
37	122	5.4	1163	9	US-09-932-257A-3	Sequence 32, Appl
38	122	5.4	2843	8	US-08-681-219-32	Sequence 1, Appl
39	121	5.3	1404	10	US-09-811-045A-1	Sequence 16, Appl
40	120.5	5.3	28	9	US-09-812-485A-37	Sequence 364, App
41	120.5	5.3	972	10	US-09-924-154-16	Sequence 1, Appl
42	120.5	5.3	1435	9	US-10-153-273-4	Sequence 4, Appl
43	120	5.3	1250	10	US-09-801-368-364	Sequence 364, App
44	120	5.3	2843	9	US-09-987-482-1	Sequence 1, Appl
45	119.5	5.2	1007	9	US-10-211-133-7	Sequence 7, Appl

ALIGNMENTS

```
RESULT 1
US-09-814-550-2
; Sequence 2, Application US/09814550
; Patent No. US20020102641A1
; GENERAL INFORMATION:
; APPLICANT: Schiavi, Susan
; APPLICANT: Madden, Stephen
; APPLICANT: Manavalan, Parthasarathy
; APPLICANT: Levine, Michael
; APPLICANT: Jan de Beur, Suzanne
; TITLE OF INVENTION: ONCOGENIC OSTEOBLAST-RELATED GENE 1
; FILE REFERENCE: 5014US
; CURRENT APPLICATION NUMBER: US/09/814,550
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US 60/191,786
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: US 60/241,598
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 525
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-814-550-2
Query Match          99.6%  Score 2269;  DB 10;  Length 525;
Best Local Similarity 99.5%  Pred. No. 8e-156;
Matches 428;  Conservative 1;  Mismatches 1;  Indels 0;  Gaps 0;
```

QY 1 VNKEYSISNKENHNGLRMSITPSTGNGKFEDGDAISKLHDEEYGAALIRNNQHIM 60
DB 96 LNKEYSISNKENHNGLRMSITPSTGNGKFEDGDAISKLHDEEYGAALIRNNQHIM 155
QY 61 GPVTAIKLGEENKENTPRNVLIIPASNNYAKAHSKOKKPPORDSOAKSPYKKSSTHR 120
DB 156 GPVTAIKLGEENKENTPRNVLIIPASNNYAKAHSKOKKPPORDSOAKSPYKKSSTHR 215
QY 121 IOHNIDYIKLHSLVKKIPSPDESGSYTDLOERGDNDISPPSGGOGPFKIDPGKGEATGPD 180
DB 216 IOHNIDYIKLHSLVKKIPSPDESGSYTDLOERGDNDISPPSGGOGPFKIDPGKGEATGPD 275

Qy	181	LEBDIOGFGEPEAESETHLDTKKPCYNELPEEEENGWNTIGRDTAEAAVAVSYL	240
Dd	276	LEGGKIDQGFGEPSAESETHLDTKKPGNEIPEKEENGWNTIGRDTAEAAVAVSYL	335
Qy	241	EGSNDIMGSTNFKELPGHEGKNRVAGSOMAHOGVEEYHYPPAPSEKEREKGGSDAAESTN	300
Dd	336	EGSNDIMGSTNFKELPGHEGKNRVAGSOMAHOGVEEYHYPPAPSEKEREKGGSDAAESTN	395
Qy	301	YNEIPIKNGKSTIRKGVDSHNSNNAOTLWDEKORFPEKGGSGGLPIPSRLDMEIKNEMDSFN	360
Dd	396	YNEIPIKNGKSTIRKGVDSHNSNNAOTLWDEKORFPEKGGSGGLPIPSRLDMEIKNEMDSFN	455
Qy	361	GPSHENIITTHRKAKHYVPHRONSTIRKMGPOGCGSGRGROPHSNRRPSSRRPDDSSSD	420
Dd	456	GPSHENIITTHRGVHYVPHRONSTIRKMGPOGCGSGRGROPHSNRRPSSRRPDDSSSD	515
Qy	421	SGSSSESDEGD 430	
Dd	516	SGSSSESDEGD 525	

```

RESULT 2
US-09-812-485A-1
; Sequence 1, Application US/09812485A
; Publication No. US20020197267A1
; GENERAL INFORMATION:
; APPLICANT: Kumagai, Yoshinari
; APPLICANT: Blacher, Russel
; APPLICANT: Yoneda, Toshiyuki
; TITLE OF INVENTION: Integlin Binding Motif Containing
; TITLE OF INVENTION: Peptides and Methods of Treating Skeletal Diseases
; FILE REFERENCE: BEAR-006CIP
; CURRENT APPLICATION NUMBER: US/09/812,485A
; CURRENT FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 09/641,034
; PRIOR FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 97
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: peptidic compound
; US-09-812-485A-1

```

[illegible]

```

RESULT 3
US-09-812-485A-2
Sequence 2, Application US/09812485A
Publication No. US20020197267A1
GENERAL INFORMATION:
APPLICANT: Kumagai, Yoshinari
APPLICANT: Kumagai, Russel
APPLICANT: Yoneda, Toshiyuki
TITLE OF INVENTION: Integritin Binding Motif Containing
TITLE OF INVENTION: Peptides and Methods of Treating Skeletal Diseases
FILE REFERENCE: BEAR-008CIP
CURRENT APPLICATION NUMBER: US/09/812,485A
CURRENT FILING DATE: 2001-03-19
PRIOR APPLICATION NUMBER: 09/641,034

```

```

, PRIOR FILING DATE: 2000-08-16
, NUMBER OF SEQ ID NOS: 50
, SOFTWARE: FastSeq for Windows Version 4.0.
, SEQ ID NO: 2
, LENGTH: 47
, TYPE: PR
, ORGANISM: Artificial Sequence
, FEATURE:
, OTHER INFORMATION: peptidic compound
US-03-812-485A-2

```

	Query Match	Similarity	10.8%	Score 246	DB 9	Length 47
Best Local	Similarity	100.0%	Pred. No. 1.8e-11			
Matches	47	Conservative	0	Mismatches	0	Indels
					0	Gaps
07	108	AKSPVKSSTRIQHNIDYLLKHLKSVKKIPSPDFECGGYTDLOERCD	154			
DB	1	AKSPVKSSTRIQHNIDYLLKHLKSVKKIPSPDFECGGYTDLOERCD	47			

```

RESULT 4 \
US-09-812-485A-4
; Sequence 4, Application US/09812485A
; Publication No. US20020197267A1
; GENERAL INFORMATION:
; APPLICANT: Kumagai, Yoshinari
; APPLICANT: Blachar, Russel
; APPLICANT: Yoneda, Toshiyuki
; TITLE OF INVENTION: Integrin Binding Motif Containing
; TITLE OF INVENTION: Peptides and Methods of Treating Skeletal Diseases
; FILE REFERENCE: BEAR-006CIP
; CURRENT APPLICATION NUMBER: US/09/812,485A
; CURRENT FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 09/641,034
; PRIOR FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 47
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: peptidic compound
US-09-812-485A-4

```

	Query Match	Similarity	Score	DB	Length
Best Local	47	94.0%	235.5	9	47
Matches	47	Conservative	0	Mismatches	0
				Indels	3
				Gaps	1

```

RESULT 5
US-09-812-485A-3
: Sequence 3, Application US/09812485A
: Publication No. US20020197267A1
: GENERAL INFORMATION:
: APPLICANT: Kumagai, Yoshihara
: APPLICANT: Blacher, Russel
: APPLICANT: Yoneda, Toshiyuki
: TITLE OF INVENTION: Integrin Binding Motif Containing
: TITLE OF INVENTION: Peptides and Methods of Treating Skeletal Diseases
: FILE REFERENCE: BEAR-006CIP
: CURRENT APPLICATION NUMBER: US/09/812,485A
: CURRENT FILING DATE: 2001-03-19
: PRIOR APPLICATION NUMBER: 09/641,034
: PRIOR FILING DATE: 2000-08-16
: NUMBER OF SEQ. ID NOS: 50
: SOFTWARE: FastSeq for Windows Version 4..0
: SEQ. ID NO. 3
: LENGTH: 47

```

TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: peptidic compound
US-09-812-485A-3

Query Match 10.0%; Score 229; DB 9; Length 47;
Best Local Similarity 100.0%; Pred. No. 3.1e-10;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 108 AOKSPVSKSTHRIQHNIDYKHLKSKVKKIPSPDESGGYTDLOE 151
4 AOKSPVSKSTHRIQHNIDYKHLKSKVKKIPSPDESGGYTDLOE 47

RESULT 6
US-09-812-485A-23
Sequence 23, Application US/09812485A
Publication No. US20020197267A1
GENERAL INFORMATION:
APPLICANT: Kumagai, Yoshinari
APPLICANT: Blacher, Russel
APPLICANT: Yoneda, Toshiyuki
TITLE OF INVENTION: Integrin Binding Motif Containing Peptides and Methods of Treating Skeletal Diseases
FILE REFERENCE: BEAR-006CIP
CURRENT APPLICATION NUMBER: US/09/812,485A
CURRENT FILING DATE: 2001-03-19
PRIOR APPLICATION NUMBER: 09/641,034
PRIOR FILING DATE: 2000-08-16
NUMBER OF SEQ ID NOS: 50
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 23
LENGTH: 40
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: peptidic compound
US-09-812-485A-23

Query Match 9.6%; Score 218; DB 9; Length 40;
Best Local Similarity 100.0%; Pred. No. 1.6e-09;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 RGNNDISPFSGDQPKKIDIPGKGATGPDLEGKDIQTGFA 191
Db 1 RGNNDISPFSGDQPKKIDIPGKGATGPDLEGKDIQTGFA 40

RESULT 7
US-09-812-485A-6
Sequence 6, Application US/09812485A
Publication No. US20020197267A1
GENERAL INFORMATION:
APPLICANT: Kumagai, Yoshinari
APPLICANT: Blacher, Russel
APPLICANT: Yoneda, Toshiyuki
TITLE OF INVENTION: Integrin Binding Motif Containing Peptides and Methods of Treating Skeletal Diseases
FILE REFERENCE: BEAR-006CIP
CURRENT APPLICATION NUMBER: US/09/812,485A
CURRENT FILING DATE: 2001-03-19
PRIOR APPLICATION NUMBER: 09/641,034
PRIOR FILING DATE: 2000-08-16
NUMBER OF SEQ ID NOS: 50
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6
LENGTH: 44
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: peptidic compound
US-09-812-485A-6

Query Match 9.5%; Score 216; DB 9; Length 44;
Best Local Similarity 88.0%; Pred. No. 2.4e-09;
Matches 44; Conservative 0; Mismatches 0; Indels 6; Gaps 1;

QY 105 DSOAKSPVSKSTHRIQHNIDYKHLKSKVKKIPSPDESGGYTDLOE 154
Db 1 DSOAKSPVSKSTHRIQHNIDYKHLKSKVKKIPSPDESGGYTDLOE 44

RESULT 8
US-09-812-485A-5
Sequence 5, Application US/09812485A
Publication No. US20020197267A1
GENERAL INFORMATION:
APPLICANT: Kumagai, Yoshinari
APPLICANT: Blacher, Russel
APPLICANT: Yoneda, Toshiyuki
TITLE OF INVENTION: Integrin Binding Motif Containing Peptides and Methods of Treating Skeletal Diseases
FILE REFERENCE: BEAR-006CIP
CURRENT APPLICATION NUMBER: US/09/812,485A
CURRENT FILING DATE: 2001-03-19
PRIOR APPLICATION NUMBER: 09/641,034
PRIOR FILING DATE: 2000-08-16
NUMBER OF SEQ ID NOS: 50
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5
LENGTH: 44
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: peptidic compound
US-09-812-485A-5

Query Match 9.4%; Score 215; DB 9; Length 44;
Best Local Similarity 100.0%; Pred. No. 2.9e-09;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 SPVSKSTHRIQHNIDYKHLKSKVKKIPSPDESGGYTDLOE 151
Db 4 SPVSKSTHRIQHNIDYKHLKSKVKKIPSPDESGGYTDLOE 44

RESULT 9
US-09-812-485A-27
Sequence 27, Application US/09812485A
Publication No. US20020197267A1
GENERAL INFORMATION:
APPLICANT: Kumagai, Yoshinari
APPLICANT: Blacher, Russel
APPLICANT: Yoneda, Toshiyuki
TITLE OF INVENTION: Integrin Binding Motif Containing Peptides and Methods of Treating Skeletal Diseases
FILE REFERENCE: BEAR-006CIP
CURRENT APPLICATION NUMBER: US/09/812,485A
CURRENT FILING DATE: 2001-03-19
PRIOR APPLICATION NUMBER: 09/641,034
PRIOR FILING DATE: 2000-08-16
NUMBER OF SEQ ID NOS: 50
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 27
LENGTH: 45
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: peptidic compound
US-09-812-485A-27

Query Match 9.4%; Score 214.5; DB 9; Length 45;
Best Local Similarity 93.3%; Pred. No. 3.2e-09;
Matches 42; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

OY 160 FSGDGPFKDIPGKGATGPDLEKDIQTGFAGPSEAE---THL 201
 DB 1 FSGDGPFKDIPGKGATGPDLEKDIQTGFAGPSEAEHSGDHL 45

RESULT 10
 US-09-812-485A-8
 ; Sequence 8, Application US/09812485A
 ; Publication No. US20020197267A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Kumagai, Yoshinari
 ; APPLICANT: Blacher, Russel
 ; APPLICANT: Yoneda, Toshiyuki
 ; TITLE OF INVENTION: Integrin Binding Motif Containing
 ; Peptides and Methods of Treating Skeletal Diseases
 ; FILE REFERENCE: BEAR-006CIP
 ; CURRENT APPLICATION NUMBER: US/09/812,485A
 ; CURRENT FILING DATE: 2001-03-19
 ; PRIOR APPLICATION NUMBER: 09/641,034
 ; PRIOR FILING DATE: 2000-08-16
 ; NUMBER OF SEQ ID NOS: 50
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 8
 ; LENGTH: 41
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Peptidic compound
 US-09-812-485A-8

Query Match 8.7%; Score 198.5; DB 9; Length 41;
 Best Local Similarity 82.0%; Pred. No. 4.1e-08;
 Matches 41; Conservative 0; Mismatches 0; Indels 9; Gaps 1;

OY 105 DSOAOKSPVSKSTHRIQHNIDYKLHLSKVKKIPSDFFSGSYTDLQERGD 154
 DB 1 DSOAOKSPVSKSTHRIQHNIDYKLHLSKVKKIPSDF-----ERGD 41

RESULT 11
 US-09-812-485A-24
 ; Sequence 24, Application US/09812485A
 ; Publication No. US20020197267A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Kumagai, Yoshinari
 ; APPLICANT: Blacher, Russel
 ; APPLICANT: Yoneda, Toshiyuki
 ; TITLE OF INVENTION: Integrin Binding Motif Containing
 ; Peptides and Methods of Treating Skeletal Diseases
 ; FILE REFERENCE: BEAR-006CIP
 ; CURRENT APPLICATION NUMBER: US/09/812,485A
 ; CURRENT FILING DATE: 2001-03-19
 ; PRIOR APPLICATION NUMBER: 09/641,034
 ; PRIOR FILING DATE: 2000-08-16
 ; NUMBER OF SEQ ID NOS: 50
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 24
 ; LENGTH: 40
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Peptidic compound
 US-09-812-485A-24

Query Match 8.4%; Score 190.5; DB 9; Length 40;
 Best Local Similarity 92.5%; Pred. No. 1.5e-07;
 Matches 37; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

OY 152 RGDNDISPFSGDGPFKDIPGKGATGPDLEKDIQTGFAGP 191
 DB 4 RGD---SPFSGDGPFKDIPGKGATGPDLEKDIQTGFAG 40

RESULT 12
 US-09-812-485A-32
 ; Sequence 32, Application US/09812485A
 ; Publication No. US20020197267A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Kumagai, Yoshinari
 ; APPLICANT: Blacher, Russel
 ; APPLICANT: Yoneda, Toshiyuki
 ; TITLE OF INVENTION: Integrin Binding Motif Containing
 ; Peptides and Methods of Treating Skeletal Diseases
 ; FILE REFERENCE: BEAR-006CIP
 ; CURRENT APPLICATION NUMBER: US/09/812,485A
 ; CURRENT FILING DATE: 2001-03-19
 ; PRIOR APPLICATION NUMBER: 09/641,034
 ; PRIOR FILING DATE: 2000-08-16
 ; NUMBER OF SEQ ID NOS: 50
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 32
 ; LENGTH: 40
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Peptidic compound
 US-09-812-485A-32

Query Match 8.3%; Score 188.5; DB 9; Length 40;
 Best Local Similarity 92.5%; Pred. No. 2.1e-07;
 Matches 37; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

OY 160 FSGDGPFKDIPGKGATG---PDLEKDIQTGFAGPSEA 196
 DB 1 FSGDGPFKDIPGKGATGPDLEKDIQTGFAGPSEA 40

RESULT 13
 US-09-812-485A-7
 ; Sequence 7, Application US/09812485A
 ; Publication No. US20020197267A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Kumagai, Yoshinari
 ; APPLICANT: Blacher, Russel
 ; APPLICANT: Yoneda, Toshiyuki
 ; TITLE OF INVENTION: Integrin Binding Motif Containing
 ; Peptides and Methods of Treating Skeletal Diseases
 ; FILE REFERENCE: BEAR-006CIP
 ; CURRENT APPLICATION NUMBER: US/09/812,485A
 ; CURRENT FILING DATE: 2001-03-19
 ; PRIOR APPLICATION NUMBER: 09/641,034
 ; PRIOR FILING DATE: 2000-08-16
 ; NUMBER OF SEQ ID NOS: 50
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 7
 ; LENGTH: 37
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Peptidic compound
 US-09-812-485A-7

Query Match 8.0%; Score 182; DB 9; Length 37;
 Best Local Similarity 100.0%; Pred. No. 5.6e-07;
 Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 118 THRIQHNIDYKLHLSKVKKIPSDFFSGSYTDLQ 151
 DB 4 THRIQHNIDYKLHLSKVKKIPSDFFSGSYTDLQ 37

RESULT 14
 US-09-812-485A-10
 ; Sequence 10, Application US/09812485A
 ; Publication No. US20020197267A1
 ; GENERAL INFORMATION:

```

; APPLICANT: Kumagai, Yoshinari
; APPLICANT: Blacher, Russel
; APPLICANT: Yoneda, Toshiyuki
; TITLE OF INVENTION: Integrin Binding Motif Containing
; TITLE OF INVENTION: Peptides and Methods of Treating Skeletal Diseases
; FILE REFERENCE: BEAR-006CIP
; CURRENT APPLICATION NUMBER: US/09/812,485A
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 09/641,034
; PRIOR FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 38
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: peptidic compound
US-09-812-485A-10

```

```

Query Match          7.9%; Score 180; DB 9; Length 38;
Best Local Similarity 76.0%; Pred. No. 8e-07;
Matches: 38; Conservative 0; Mismatches 0; Indels 12; Gaps 1;

```

```

QY 105 DSOAQKSPVKSSTHRIQHNIDYKLSKVKKIPSDFEKSGYTDLERGD 154
DB 1 DSOAQKSPVKSSTHRIQHNIDYKLSKVKKIPSDFEKSGYTDLERGD 154

```

```

RESULT 15
US-09-812-485A-25
; Sequence 25, Application US/09812485A
; Publication No. US20020197267A1
; GENERAL INFORMATION:
; APPLICANT: Kumagai, Yoshinari
; APPLICANT: Blacher, Russel
; APPLICANT: Yoneda, Toshiyuki
; TITLE OF INVENTION: Integrin Binding Motif Containing
; TITLE OF INVENTION: Peptides and Methods of Treating Skeletal Diseases
; FILE REFERENCE: BEAR-006CIP
; CURRENT APPLICATION NUMBER: US/09/812,485A
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 09/641,034
; PRIOR FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 25
; LENGTH: 35
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: peptidic compound
US-09-812-485A-25

```

```

Query Match          7.2%; Score 163.5; DB 9; Length 35;
Best Local Similarity 91.4%; Pred. No. 1.1e-05;
Matches: 32; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

```

```

QY 155 NDISPF---SGDQPFKDIKGEATGPDLEKDI 186
DB 1 NDISPFKDSGDQPFKDIKGEATGPDLEKDI 35

```

Search completed: April 11, 2003, 12:04:30
Job time : 21 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: April 16, 2003, 19:03:35 ; Search time 3826 Seconds
(without alignments)
3270.833 Million cell updates/sec

Title: US-09-700-696B-2
Perfect score: 2279
Sequence: 1 VNKEYISKNKENTHNGLRMS.....RRDSESSSDSSSESDDG 430

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 segs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL-frame+ p2n.model -DEV-xlh
-O/cgn2.1/USPTO.spool/US09700696/runat.11042003.101724.12384/app.query.fasta.1.583
-DB-GenEmbl -QFMT-fastap -SUFFIX-p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS-bits -START=1 -END=1 -MATRIX-blosum62 -TRANS-human40.cdi -LIST=45
-DOCALIGN=200 -THR-score-pct -THR-MAX=100 -THR-MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT-pco NOR-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09700696.ecgn.1.1.2436-@runat.11042003.101724.12384 -NCPu=6 -ICPU=3
-NO_XLPEX -NO_MMAP -LARGEQUERY -NEG-SCORES=0 -WAIT -LONGLOG -DEV-TIMEOUT=120
-XGAPOP=10 -YGAPEXT=0.5 -DELop=6 -DELEXt=7

Database : GenEmbl: *
1: gb_da: *
2: gb_htg: *
3: gb_in: *
4: gb_ov: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *
15: em_da: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_to: *
27: em_sts: *
28: em_un: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2279	100.0	1655	6 AX010293	AX010293 Sequence
2	2279	100.0	1655	6 AX135682	AX135682 Sequence
3	2276	99.9	1617	9 AF325916	AF325916 Homo sapi
4	2276	99.9	1989	9 HSA276396	AJ276396 Homo sapi
5	2276	99.9	2013	6 AX135707	AX135707 Sequence
6	2276	99.9	187624	9 AC093768	AC093768 Homo sapi
7	2058.5	90.3	2095	9 AB046056	AB046056 Macaca fa
8	2058.5	90.3	2259	9 AB050891	AB050891 Macaca fa
9	2054.5	90.1	2140	9 AB050259	AB050259 Macaca fa
10	2031.5	89.1	2085	9 AB056814	AB056814 Macaca fa
11	983.5	43.2	1580	10 AF314964	AF314964 Mus muscu
12	983.5	43.2	1682	10 AF228661	AF228661 Mus muscu
13	983.5	41.9	169603	2 AC129695	AC129695 Dictyoste
14	938.5	41.2	1651	10 AF530559	AF530559 Rattus no
15	938.5	41.2	1672	10 AF530558	AF530558 Rattus no
16	938.5	41.2	1672	10 AF530559	AF530559 Rattus no
17	418.5	18.4	60588	2 AC124106	AC124106 Mus muscu
18	207.5	9.1	171187	2 AC116960	AC116960 Dictyoste
19	182	8.0	5888	3 AF230482	AF230482 Hydra vtl
20	165	7.2	5145	2 AC015315	AC015315 Drosophi
21	165	7.2	303760	3 AC012373	AC012373 Drosophi
22	165	7.1	83881	8 AP004917	AP004917 Rattus no
23	161	6.9	3394	10 AF247187	AF247187 Rattus no
24	157	6.9	4440	10 AF250374	AF250374 Rattus no
25	157	6.9	4457	10 RNC0403971	RNC0403971 Rattus no
26	157	6.9	10029	1 AE006040	AE006040 Pasteurel
27	156.5	6.8	2982	8 AF476961	AF476961 Aegilops
28	155.5	6.8	182987	2 AL772238	AL772238 Mus muscu
29	155.5	6.8	2391	8 AF476959	AF476959 Aegilops
30	154	6.8	182525	3 AC009213	AC009213 Drosophi
31	153	6.6	3023	4 BTU47636	BTU47636 Bos taur
32	151.5	6.6	129825	6 AC128257	AC128257 Rattus no
33	151.5	6.6	4814	4 CFA296287	CFA296287 Canis fam
34	151	6.6	8490	5 XLU06442	XLU06442 Xenopus lae
35	150	6.6	8646	2 AC014412	AC014412 Drosophi
36	150	6.6	154201	2 AC105627	AC105627 Rattus no
37	150	6.6	239171	3 AE003765	AE003765 Drosophi
38	149.5	6.6	156060	2 AC004153	AC004153 Plasmodiu
39	149.5	6.6	236406	2 AC079489	AC079489 Mus muscu
40	149	6.5	83549	2 AL356504	AL356504 Human DNA
41	149	6.5	176354	2 AC111568	AC111568 Rattus no
42	148.5	6.5	9944	9 AF163151	AF163151 Homo sapi
43	148.5	6.5	189286	9 AC093906	AC093906 Homo sapi
44	148.5	6.5	3045	9 AK096526	AK096526 Homo sapi
45	148	6.5	3045	9 AK096526	AK096526 Homo sapi

RESULT 1

ALIGNMENTS

AX010293 1655 bp DNA Linear PAT 06-SEP-2000
 LOCUS AX010293
 DEFINITION Sequence 1 from Patent WO9960017.
 ACCSSION AX010293
 VERSION AX010293.1 GI:9997173
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE
 AUTHORS Rowe, P.
 TITLE A novel polypeptide hormone phosphatonin
 JOURNAL Patent: WO 9960017-A 1 25-NOV-1999;
 UNITV LONDON (GB); ROWE, PETER (GB)
 FEATURES
 source Location/Qualifiers
 1..1655
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 <1..1293
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAC07513.1"
 /db_xref="GI:9997174"
 /translation="VAKESISKENTHNGLMSTYPTKNGKGEEDGDAISKLHDO
 ERYGALIRNNQIHMPVYAIKLGEENKENTPRNVLTIPASMYAKASKDKKK
 ORDSQKSPKSKSTHRIQHNIDYLLKHSKYKKLPSDEGGGYDLDQERNDISP
 SGDGQFKEIDPGKGEATGPDLEKIDQTFAGSESEHLDTKRGVNEIDEREENG
 GNTGIRDETAKEADADVSLVSGEINFTNFKLPREBNRYVAGSOMAHQGVVE
 FHYPRAPEKREKSSDAEESTNNEITRKNGKSGTRKVDNSNRQATLNEKORFPS
 KKSQGLPIPSKRLDNEIKENMDSFNGSPHEINTTHRGYHYVPHNQNSTNRKMPQ
 GKSGWGRPHSPNFRFSRRDSSSDSGSSSDSGD"
 BASE COUNT 609 a 303 c 380 g 363 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1 19e-122 Length: 1655
 Score: 2279.00 Matches: 430
 Percent Similarity: 100.008 Conservative: 0
 Best Local Similarity: 100.008 Mismatches: 0
 Query Match: 100.004 Indels: 0
 DB: Gaps: 0
 US-09-700-696b-2 (1-430) x AX010293 (1-1655)
 QY 1 Valasnysglutyrserlleserasnlysluasnthrhisaenglyleuargmetser 20
 DB 1 GTGAATAAAGAAATATAGATACGTAACAAAGAAATATCTCAATGGCTGAGATGCA 60
 QY 21 lletyrprolysserthglyasnlysglyphegluaspglyaspapalaillserlys 40
 DB 61 ATTTATCTTAAGTCACTACGGAATTAAGGATTGAGATGAGATGATGATGACGAAA 120
 QY 41 leuhtisapglngluglutyrglyalaalaleuileargasnmetglhislmet 60
 DB 121 CTACATGACCAAGAAGATATGGCCACCTCTCATCAGAAATATCAATCAATATAATG 180
 QY 61 glyprovalthratallelyslleuenglygluasnylsgluasnthrhproargasn 80
 DB 181 GGGCCAGAGACGCGATTAACTCTGGGGGAGAAACAAAGAAACACACCTTAGGAT 240
 QY 81 valleunsllelleproalasermetasntratalatylasahislerlysaplysls 100
 DB 241 GTTCTTAACATATATCCCAAGCAAGTATGATTAAGCAACACTCAAGAGATTAAGA 300
 QY 101 lysproglinaraspsercinalacnlysserprovallysserlyserthrhysarg 120
 DB 301 AAGCCCTCAAGAGATTTCCCAAGCCAGAAAGCTCCAGTAAAAAGCAAGCACTGCT 360
 QY 121 lileglnhisasnllasprylleuylshisleserlyslvallyslleproserasp 140
 DB 361 ATTCAACACACAACTTACTTAACATCTCTCAAAAAGTCAAAAAAATATCCCACTGAT 420

QY 141 phegluglyserglytythrthraspleuenglualarglyaspasnaspleserprophe 160
 DB 421 TTTGAAGCAGCGGCTTATACAGATCTTCAAGAGAGAGGGCAATGATATATCTCTTC 480
 QY 161 serglyaspgllyglprophelyaspilleproglylysglyglualaathrhglyproasp 180
 DB 481 AGTGGGAGCGGCCACCTTTTAAGSACATCTCTGTTAAGAGAGAACTACTGCTCGAC 540
 QY 181 leugluglylyaspilleglthrthrhphlealaglyproserglualaaglyserthhis 200
 DB 541 CTAGAAGCAAGATATTCAAACAGGTTTGCAGGCCCAAGTGAAGGAGTGAAGTCTAT 600
 QY 201 leuasprthlyslsproglytyrasngluilleprogluargluasnglylyasn 220
 DB 601 CTTGACACAAAAGAGCCAGTTATATGATGATCCAGAGAGAGAAATAGTGAAT 660
 QY 221 thrileglythrargasppluThrthralasglualaasapalaavalasylserleuval 240
 DB 661 ACCATTGGAAGTAAAGGATGAACCTCGAAGAGCCAGATGCTGTGATGTCAGCTTGTGA 720
 QY 241 gluglyserasnaspillemetglyserthrthrasphelysgluileuproglyargglugly 260
 DB 721 GAGGCGAGCAAGATATCATGAGTGTAGTACCAATTTTAAGAGCTCTCGTGAAGAGAGA 780
 QY 261 asnarvalaspalaaglysercinsasnaahisglinglylyslvalgluphehistrypro 280
 DB 781 AACAGATGATGCTGCGAGCCAAATGCTCCAAAGGAGAGTGAAGTTCATTATTCCT 840
 QY 281 proalaproserlysgluysarglysglyserasnaspalalaglyserthrasn 300
 DB 841 CCGCACCCCTCAAAAGAAAGAAAGAAAGAGGAGTGTGATGACCTAAAGTACCAAC 900
 QY 301 tyrasngluilleprolysaenglylysglyserthrarglyslvalasphisserasn 320
 DB 901 TATATGAATTCCTTAATAATGCAAGGAGACACCAAGAGGTTAGATCATCTTAAT 960
 QY 321 argasnlnalathrhleuasnglysluysargphproserlysglysserengly 340
 DB 961 AGGACCAACCAACCTTAATGAAAAAAGGTTCTTGTAGAGGCCAAAGTACGGGC 1020
 QY 341 leuprothleproserargglyleuasasnngluillelyasnnglumetaspserpheasn 360
 DB 1021 CTGCCATTCCTCTCTGCTGCTGATTAATGAATCAAAAACCAATGATCTCTTAAT 1080
 QY 361 glyproserhisgluasnillethrthrhisglyarglyslthrhistryvalprohisarg 380
 DB 1081 GGGCCGATCATGAGATATATATACACATGCAAGAAATATATATATATATATATATAT 1140
 QY 381 gluasnsnserthrargasnlysglymetproglnglylysglysertrpglyarglin 400
 DB 1141 CAAATATATCTACACGGAATTAAGGATATATCCACAAAGGAAAGGCTCTGGGTAGACAA 1200
 QY 401 prohisserasnargargpheaseraserargargaspserasergluseraserasp 420
 DB 1201 CCCCATTCCAACAGAGAGTATGTTCCCTAGAAAGGATGACAGTATGATCATCTGAC 1260
 QY 421 serglyserasergluserasp 430
 DB 1261 AGTGGCAGTTCAAGTGAAGCGATGGAC 1290
 RESULT 2
 AX135682 1655 bp DNA Linear PAT 29-MAY-2001
 LOCUS AX135682
 DEFINITION Sequence 1 from Patent WO0132878.
 ACCSSION AX135682
 VERSION AX135682.1 GI:14271952
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE
 AUTHORS Rowe, P.

TITLE Polypeptide hormone phosphatoin
JOURNAL Patent: WO 0132878-A 1 10-MAY-2001;
UNIVERSITY COLLEGE LONDON (GB)
FEATURES Location/Qualifiers
source 1..1655
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 609 a 303 c 380 g 363 t
ORIGIN

Alignment Scores:

Pred. No.: 1,19e-122 Length: 1655
Score: 2279.00 Matches: 430
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-700-696b-2 (1-430) x AX135682 (1-1655)

QY 1 ValAsnLysGluThrSerIleSerAsnLysGluAsnThrHisAsnGlyLeuArgMetSer 20
DB 1 GTGAATTAAGCAATTAATCTACTAACAAGAAATTAATCTCAATGCGCTGAGAGATGCA 60
QY 21 IleTy-ProLysSerThrGlyAsnLysGlyPheGluAspGlyAspAlaIleSerLys 40
DB 61 ATTATTCCTAAGTCACTGGAGTAATAAGGTTTGAGAGATGAGATGCTATCAGCAAA 120
QY 41 LeuHisAspGlnGluGlyArgAlaLeuIleArgAsnAsnMetGlnHisIleMet 60
DB 121 CTTCATGACCCAAAGAAAGATATGGCGCGCTCTCATGAAATTAACATGCAACATATATG 180
QY 61 GlyProValThrAlaIleLysLeuLysGluGluAsnLysGluAsnThrProArgAsn 80
DB 181 GGGCCAGTACTCGATTAACTCTGGGGAAGAAACAAAGAAACAACTAGGAAT 240
QY 81 ValLeuAsnIleIleProAlaSerMetAsnTyraIleLysAlaHisSerLysAspLysLys 100
DB 241 GTTCATAACATTAATCCAGCAAGATGAATTAATGCTAAAGCACACTGGAATATAAAG 300
QY 101 LysProGlnArgAspSerGlnAlaGlnLysSerProValLysSerLysSerThrHisArg 120
DB 301 AACCTCTAAAGATATCCCAAGCCCAAGAAAGTCAGTAAAGCAAAACACCCATCGT 360
QY 121 IleGlnHisAsnIleAspTyraLysLysHisLeuSerLysValLysLysIleProSerAsp 140
DB 361 ATTCAACACAACTGACTCTCTTAAACATCTCTCAAAAGTCAAAAAATCCCAAGTAT 420
QY 141 PheGluGlySerGlyTyraThrAspLeuGlnLysArgGlyAspAsnAspIleSerProPhe 160
DB 421 TTGGAAGGCGAGCGTTATACAGATCTCAAGAGAGAGGCGCAATGATATATCTCTTC 480
QY 161 SerGlyAspGlyGlnProPheLysAspIleProGlyLysGlyLysAlaThrGlyProAsp 180
DB 481 AGTGGGAGCGCAACCTTTTAAGGCAATTCCTGTAAAGAGAAAGCAAGCTAGTCCGAC 540
QY 181 LeuGlnGlyLysAspIleGlnThrGlyPheAlaGlyProSerGlnAlaGlnSerThrHis 200
DB 541 CTGAAGGCAAAATATTCAAACAGGTTTGCAAGGCCCAAGTGAAGTGAAGTACTCAT 600
QY 201 LeuAspThrLysLysProGlyTyraGlnLysIleProGluArgGlnLysAsnGlyLysAsn 220
DB 601 CTTGACACAAAGAAAGCCAGCTTATTAATGAGATCCCAAGAGAGAAATGCTGCAAT 660
QY 221 ThrIleGlyThrArgAspGluThrAlaLysGluAlaAspAlaValAspValSerLeuVal 240
DB 661 ACCATTGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 720
QY 241 GluGlySerAsnAspIleMetGlySerThrAsnPheLysGluLeuProGlyArgGlnGly 260
DB 721 GAGGCGCAAGCATATTCATGAGGTAGTACCAATTTTAAGAGAGCTCCCTGGAAGAAAGA 780
QY 261 AsnArgValAspAlaGlySerGlnAsnAlaHisGlnGlyLysValGluPheHisTyraPro 280

DB 781 AACAGATGATCTGCGACCAAAATGCTCACCAAGGGAAGTTGCTTACATTAACCT 840
QY 281 ProAlaProSerLysGluLysArgLysGluGlySerSerAspAlaIleGlnSerThrAsn 300
DB 841 CCTGCACCTTCACAAAGCAAGAAAGAAAGAGAGAGTATGATGAGCTGAAGTACCAC 900
QY 301 TyraGlnIleProLysAsnGlyLysGlySerThrArgLysGlyAlaAspHisSerAsn 320
DB 901 TATAATGAATTCCTTAATAATGCAAGGAGTACAGCAAGAAAGGTTGATCATTTCTAAT 960
QY 321 ArgAsnGlnAlaThrLeuAsnGluLysGlnArgPheProSerLysGlyLysSerGlnGly 340
DB 961 AGGAAACCAACCAACCTTAATTAAGAAAAAAGGTTTCTCTAAGGCAAAAGTCAAGGCG 1020
QY 341 LeuProIleProSerArgGlyLeuAspAsnGluIleLysAsnGlnMetAspSerPheAsn 360
DB 1021 CTGCCATTCCTTCCTGCTGATTAATGAATCAAAACGAATGAGTTCCTTTAAT 1080
QY 361 GlyProSerHisGluAsnIleIleThrHisGlyArgLysTyraIleProHisArg 380
DB 1081 GGGCCAGTATGAGATTAATTAACATGCGCAAGAAATATCATTTATGATCCCAAGCA 1140
QY 381 GlnAsnAsnSerThrArgAsnLysGlyMetProGlnGlyLysGlySerTrpGlyArgGln 400
DB 1141 CAATATTAATCTACACGGAATTAAGGTATGCCCAAGGAAAGGCTCTGGGATAGCAA 1200
QY 401 ProHisSerAsnArgArgPheSerSerArgArgAspAspSerSerGlnSerSerAsp 420
DB 1201 CCCCATTCACAAAGAGAGTTTACCTCCCTGAAGGAGTGAAGTGAAGTGAAGTGAAGTGA 1260
QY 421 SerGlySerSerSerGlnSerAspGlyAsp 430
DB 1261 AGTGGCAGTTCAAGTGAAGCATGTGAC 1290

RESULT 3 AF325916 1617 bp mRNA linear PRI 03-JUL-2001

LOCUS AF325916 Homo sapiens matrix extracellular phosphoglycoprotein precursor

DEFINITION (MEP) mRNA, complete cds.

ACCESSION AF325916 GI:14586740

VERSION AF325916

KEYWORDS

SOURCE

ORGANISM

Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 1617)
Argiro, L., Desbarats, M., Glorieux, F.H. and Ecarot, B.

Mepe, the gene encoding a tumor-secreted protein in oncogenic hypophosphatemic osteomalacia, is expressed in bone

Genomics 74 (3), 342-351 (2001)

JOURNAL

MEDLINE

PUBMED

11414762

REFERENCE

AUTHORS

TITLE

JOURNAL

Submitted (04-DEC-2000) Genetics, Shriners Hospital, 1529 Cedar

Avenue, Montreal, QC H3G 1A6, Canada

Location/Qualifiers

1..1617

/organism="Homo sapiens"

/db_xref="taxon:9606"

/cell_type="bone"

1..1617

/gene="MEPE"

40..1617

/gene="MEPE"

/codon_start=1

/product="matrix extracellular phosphoglycoprotein precursor"

/protein_id="AAK70343.1"

/db_xref="GI:14586741"

/translation="MRYFCVGLLFTVMAAPFPOPEKTQSCVEEROEKNKN
IGFHLKGRINQSTSEKENTVOERKDKDISSEASENGSSKSONVETNROKLNKYSI
SNKENTHGRIMSTYKSTGKPEDDDATSKHDEEYGAALIRNMORLNKPVTA
IKLGEENKENTPNNVNLITPASMNTYAKAHKKKQROBQAKSPYKSKSTRIORH
NIDYKHLKSVKKLIPDSQESGYTDLOERGNDSIPDSQDPRKDIPEGKEATGPDL
EKGDIQTFAGPSPAESSTHLDTKKPGVNEIPEPENGNDIGTDEDEAKEDAVSL
VEGENDIMGSTNEFELPGREGRNADAGSONAHQKVEFHYPPASKEKRGSDAE
STNNEIPKNGKSTRKGVDSHNRQATLNKORPSPKSGOGLPIPSRGLDNEIKNE
MDSFNSPHEHITTHGRKHYVPHRKNSTNNKMPQKSGWGNPSPHNSRRRD
DSSSDSGSSSESDGD
40. .87
/gene="MEPE"
mat_peptide 88.1614
/gene="MEPE"
/product="matrix extracellular phosphoglycoprotein"

BASE COUNT 604 a 300 c 373 g 340 t
ORIGIN

Alignment Scores:

Prod. No.: 1.73e-122 Length: 1617
Score: 2276.00 Matches: 429
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.77% Mismatches: 0
Query Match: 99.87% Indels: 0
DB: Gaps: 0

US-09-700-696b-2 (1-430) x AF325916 (1-1617)

QY 1 ValAsnLysGluTyrSerIleSerAsnLysGluAsnThrHisAsnGlyLeuArgMetSer 20
Db 325 CTGAATTAAGATATAGATCATGTAACAAGAGAAATGACTCAAAAGGCTGAGATGTCA 384
QY 21 IleTyrProLysSerThrGlyAsnLysGlyPheGluAsnGlyAspAlaIleSerLys 40
Db 385 ATTATCTTAAGTCACTACGGATTAAGGGTTTGAGGATGCGATGATGCTATCGCANA 444
QY 41 LeuHisAspGlnGluGluTyrGlyAlaAlaLeuIleArgAsnMetGlnHisIleMet 60
Db 445 CTACATGACCAAGAAGAAATATGGCCAGCTCTCATCAGAAATATACATGACATATATATG 504
QY 61 GlyProValThrAlaIleLysLeuLysGluGluAsnLysGluAsnThrProArgAsn 80
Db 505 GGGCCAGTACGCTGATTAACCTCTGGGGGAGAAACAAAGAGAAACACACCTGAGAT 564
QY 81 ValLeuAsnIleIleProAlaSerMetAsnTyrAlaLysAlaHisSerLysAspLysLys 100
Db 565 GTTCTAAACATATATCCCGCAAGTATGATATGCTTAAAGCACCTCGAAGATTAAGAAG 624
QY 101 LysProGlnArgAspSerGlnAlaGlnLysSerProValLysSerLysSerThrHisArg 120
Db 625 AAGCCTCAAAAGATATCCCAAGCCCAAGAAAGTCCAGTAAAGCAAAAGCACCATCGT 684
QY 121 IleGlnHisAsnIleAspTyrLeuLysHisLysSerLysValLysLysIleProSerAsp 140
Db 685 ATTCAACCAACATGACTACTTAAACATCTCTCAAAAGTCAAAAGTCCCCAGTAT 744
QY 141 PheGluGlySerGlyTyrThrAspLeuGlnGluArgGlyAspAsnAspIleSerProPhe 160
Db 745 TTTGAGGACGCGTTATACAGATTTCAAGAGAGAGGGGACATGATATATCTCTTTC 804
QY 161 SerGlyAspGlyGlnProPheLysAspIleProGlyLysGlyGluAlaThrGlyProAsp 180
Db 805 AGTGGGAGCGGCAACCTTTAAGACATCTCTGTTAAGGAGAAAGTACGCTCTTAC 864
QY 181 LeuGluGlyLysAspIleGlnThrGlyPheAlaGlyProSerGluAlaGluSerThrHis 200
Db 865 CTAGAGGCAAGATATCAAAAGGGTTTCAGGCGCCCAAGTGAAGCTGAGAGTACTCAT 924
QY 201 LeuAspThrLysLysProGlyTyrAsnGlnIleProGluLysGluGluAsnGlyLysAsn 220
Db 925 CTGGACCAAAAAGCCAGGTTATATAGATCCAGAGAGAGAAAGTGGTGAAT 984
QY 221 ThrIleGlyThrArgAspGluThrAlaLysGluAlaAspAlaValAspValSerLeuVal 240

Db 985 ACCATTGGAGACTAGGAGTGAAGAACTCGCAAGAGCAGAGATGCTGTGATGTACGCTTGA 1044
QY 241 GluGlySerAsnAspIleMetGlySerThrAsnPheLysGluLeuProGlyArgGluGly 260
Db 1045 GAGGGCACCACCAAGATATATAGGGTGTACCAATTTTAAAGAGACTCCCTGGAAGACAAGA 1104
QY 261 AsnArgValAspAlaGlySerGlnAsnAlaHisGlnGlyLysValGluPheHisTyrPro 280
Db 1105 AACACAGTGGATGCTGGCAGCCAAATGCTCCACCAAGGAGAGTTGAGTTTCATTAACCT 1164
QY 281 ProAlaProSerLysGluLysArgLysGluGlySerSerAspAlaIleGluSerThrAsn 300
Db 1165 CCGCACCCCTCAAAAGAGAAAAGAAAGAGCAGTACGATGATGACGCTGAAGTACCAAC 1224
QY 301 TyrAsnGluIleProLysAsnGlyLysGlySerThrArgLysGlyValAspHisSerAsn 320
Db 1225 TATATATGAATATCTTAAAAATGGCAAGGCACTACCAAGAGGTGTATGATCATTTCTAAT 1284
QY 321 ArgAsnGlnAlaThrLeuAsnGluLysGlnArgPheProSerLysGlyLysSerGlnGly 340
Db 1285 AGGAACCAAGCAGCACTTAATTAATAAACAAGTTTCTAGTAAGGCAAAAGTCAGGGC 1344
QY 341 LeuProIleProSerArgGlyLeuAspAsnGluIleLysAsnGluMetAspSerPheAsn 360
Db 1345 CTGCCATTCCTTCGTGCTGTGATGATTAATCAAAACAAAGATGATTCCTTTAAT 1404
QY 361 GlyProSerHisGluAsnIleIleThrHisGlyArgLysTyrHisTyrValProHisArg 380
Db 1405 GGGCCAGTCAAGATATATATTAACATGAGCAAAAATATCATATATGATACCCACAGA 1464
QY 381 GlnAsnAsnSerThrArgAsnLysGlyMetProGlnGlyLysGlySerTyrPylArgGln 400
Db 1465 CAAATATATTTCTACACGGATATAGGATGCCACAAGGAAAGGCTCTGGGGGTAGACA 1524
QY 401 ProHisSerAsnArgArgPheSerSerArgArgArgAspAspSerSerGluSerSerAsp 420
Db 1525 CCCCATTCCAAACAGAGAGGTTTGTCCCTGAGAAAGGATGACAGTACATCATCTGAC 1584
QY 421 SerGlySerSerSerGluSerAspGlyAsp 430
Db 1585 AGTGGCAGTCAAGTGAAGAGCGATGTGAC 1614
RESULT 4
HSA276396 1989 bp mRNA linear PRI 14-JUL-2000
LOCUS HSA276396
DEFINITION Homo sapiens mRNA for matrix extracellular phosphoglycoprotein
(MEPE gene).
ACCESSION AJ276396
VERSION AJ276396.1 GI:367209
KEYWORDS matrix extracellular phosphoglycoprotein; MEPE gene; osteomalacia.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 1989)
AUTHORS Rowe, P.S., de Zoyza, P.A., Dong, R., Wang, H.R., White, K.E.,
Econs, M.J. and Oudet, C.L.
TITLE MEPE, a new gene expressed in bone marrow and tumors causing
osteomalacia
JOURNAL Genomics 67 (1), 54-68 (2000)
MEDLINE 20399567
PUBMED 10945470
REFERENCE 2 (bases 1 to 1989)
AUTHORS Rowe, P.S.N.
TITLE Direct Submission
JOURNAL Submitted (06-MAR-2000) Rowe P.S.N., Biochemistry and Molecular
Biology, Royal Free and University College London Medical School,
Rowland Hill Street, Hampstead, London NW3 2PF, UK
FEATURES
Location/Qualifiers
1..1989
/organism="Homo sapiens"
/db_xref="taxon:9606"

```

gene      /chromosome="4"
          /map="4q21.1"
          /tissue_type="bone"
          50..1954
          /gene="MEPE"
          50..1627
          /gene="MEPE"
          /note="secreted from oncogenic hypophosphataemic tumours"
          /codon_start=1
          /product="matrix extracellular phosphoglycoprotein"
          /protein_id="CA97250.1"
          /db_xref="GI:9367210"
          /translacion="MRFVGVGLLFVYMAAPFPQTEKTKOSVCEEOEHOEKNKDN
          IGFHILGRINQELSSKENIVOEKDLSEASENKSSEKSONYFTNRRLNKEYSI
          SKEENTHNGRLMSIYPKSTGNKEFEDDDAISLHDOEEYGAALIRNMOMHICPTA
          IKLGEENKENTPKNVLIIPASMYAKAKSKKKRPPROSOAKSPYAKSTHRIQI
          NIDYIKLSKVKRKPSDFEGSYTDLOERDNDISFGSGQFPKIDPGKEAVTGDIL
          EGKDIOFGAPGSEASESTHDTKPKGTNELPEREENGNTIGTREDTAKRADVDVSL
          VEGNDIMGSTNFKELPREGNVDAGSQAHQKVEFHYPPAPSEKREKSGSDAAE
          STNINELPKNGKSTRGVDSNRNOATLNKORFSPKSGQGLIPSRGLDNEIKNE
          MDSFNGPSHENITTHGRKHYHPHRQNNSTRKNGMPOGKSGMGRPHSNRRFSRRRD
          DSSSSSDSGSSSSSDGD"
          50..100
          /gene="MEPE"
          101..1625
          /gene="MEPE"
          /product="matrix extracellular phosphoglycoprotein"
          1949..1954
          /gene="MEPE"
          /evidence="not experimental"
          360 c 456 g 437 t

BASE COUNT 736 a 360 c 456 g 437 t
ORIGIN

Alignment Scores:
Pred. No.: 2,18e-122 Length: 1989
Score: 2276.00 Matches: 429
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.77% Mismatches: 0
Query Match: 99.87% Indels: 0
DB: 9 Gaps: 0

US-09-700-696b-2 (1-430) x HSA276396 (1-1989)
QY 1 ValAsnLysGluTyrSerIleSerAsnLysGluAsnThrHisnGlyLeuArgMetSer 20
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 335 CTGAATAAAGAAATATAGTATACATAACAAAGAAATACATGAGCCGAGAGATGA 394
QY 21 IleTyrProLysSerThrGlyAsnLysGlyPheGluAspGlyAspAspAlaIleSerLys 40
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 395 ATTTATCTCTAAGTCAACTGGGAAATTAAGGGTTGAGGATGAGATGCTATCAACAA 454
QY 41 LeuHisAspGlnGluGluTyrGlyAlaAlaLeuIleArgAsnAsnMetGlnHisIleMet 60
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 455 CTGATGACCAACAAAGAAATATGGCGACGCTCATCAGAAATACATGCAACATATAAG 514
QY 61 GlyProValThrAlaIleLysLeuLysGlyGluGluAsnLysGluAsnThrProArgAsn 80
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 515 GGGCCAGTACGTCGATTAACTCTGGGGGAAAGAAAGAAAGAACACACCTAGGAAT 574
QY 81 ValLeuAsnIleIleProAlaSerMetAsnTyrAlaLysAlaHisSerLysAspLysLys 100
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 575 GTTCTAACAATAATCCACCAAGATATGCTTAAAGCAACACCTGCAAGATATAAG 634
QY 101 LysProGlnArgAspSerGlnAlaGlnLysSerProValLysSerLysSerThrHisArg 120
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 635 AAGCCCTCAAGAGATCCCAAGCCCAAGAAAGTCCAGTAAAGCAAAAGCAACCATCGT 694
QY 121 IleGlnHisAsnIleAspTyrLeuLysHisLeuSerLysValLysLysIleProSerAsp 140
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 695 ATTCAACACACATGACATCTAAACATCTCTCAAAAGTCAAAAGTATCCCAAGTAT 754
QY 141 PheGluGlySerGlyTyrThrAspLeuGlnArgGlyAspAsnAspLysSerProPhe 160
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::

```

```

Db 755 TTGAAGGCAAGCGTTATACAGATCTTCAAGAGAGAGCGGCAATGATATATCTCTTC 814
QY 161 SerGlyAspGlyGlnProPheLysAspIleProGlyLysGlyLysLysLysLysProAsp 180
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 815 AGTGGGAGCGGCAACCTTTTAAGACATTCCTGGTAAAGAGAAACCTGCTGCTGAC 874
QY 181 LeuGluGlyLysAspIleGlnThrGlyPheAlaGlyProSerGlnAlaLysSerThrHis 200
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 875 CTAGAAAGCAAAATATTTCAACAGGCTTTGACGGCCCAAGTGAAGCTGAGTACTCAT 934
QY 201 LeuAspThrLysLysProGlyTyrAsnGluIleProGluArgGluLysnGlyLysAsn 220
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 935 CTTGACACCAAAAAAGCCAGTTATATATGAGATCCAGAGAGAAAGAAATATGTGGAAT 994
QY 221 ThrIleGlyThrArgAspGluThrAlaLysGluAlaAspAlaValAspValSerLeuVal 240
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 995 ACCATTGGAACATGAGATGAACCTGGCAAGAGGCAAGATGCTGTGATCTCAGCCTTGA 1054
QY 241 GluGlySerAsnAspIleMetGlySerThrAsnPheLysGluLeuProGlyArgGluGly 260
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 1055 GAGGCGACAGCATATCATGCTAGTACCAATTTTAAAGACCTCCTGGAAGAGAAAGA 1114
QY 261 AsnArgValAspAlaGlySerGlnAsnAlaHisGlnGlyLysValGluPheHisTyrPro 280
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 1115 AACAGATGAGATGCTGCGACGCCAAATGCTCACCAAGGAAAGATGAGTTGATTACCTT 1174
QY 281 ProAlaProSerLysGluLysArgLysGlyLysSerSerAspAlaAlaGluSerThrAsn 300
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 1175 CCGCACCTCTCAAAAGAGAAAGAAAGAAAGAGACAGTAGTAGAGCTGAAGTCAACAA 1234
QY 301 TyrAsnGluIleProLysAsnGlyLysGlySerThrArgLysGlyValAspHisSerAsn 320
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 1235 TATATGAATTCCTTAAATATGCAAGGACATACAGAAAGGAGTGTGATCATTCATTAAT 1294
QY 321 ArgAsnGlnAlaIleThrLeuAsnGluLysGlnArgPheProSerLysGlyLysSerGlnGly 340
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 1295 AGGAACCAAGCAACCTTAAATGAAGAAACAAAGCTTCTAGTAAAGGCAAAAGTACGGC 1354
QY 341 LeuProIleProSerArgGlyLeuAspAsnGluIleLysAsnGluMetAspSerPheAsn 360
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 1355 CTGCCCATTCCTCTGCGGCTTGTATATAGAAACAAACAAAGTATCCCTTAAAT 1414
QY 361 GlyProSerHisGluAsnIleIleThrHisGlyArgLysTyrHisTyrValProHisArg 380
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 1415 GCCCCAGACATGAGATATATATACATGACAGAAATATCATATATGATCCCCACAGA 1474
QY 381 GlnAsnAsnSerThrArgAsnLysGlyMetProGlnGlyLysGlySerTyrGlyArgGln 400
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 1475 CAAATTAATTTCTACACGGGAATTAAGGTATGCCACAAAGGAAAGGCTCCTCGGTAAGCAA 1534
QY 401 ProHisSerAsnArgArgPheSerSerArgArgArgAspAspSerSerGluSerSerAsp 420
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 1535 CCCCATTCACACAGAGAGGTTTAACTTCCCGTGAAGGAGATGACATGATGATCTGAC 1594
QY 421 SerGlySerSerSerGluSerAspGlyAsp 430
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 1595 AGTGCACTTCAAGTAGAGAGCATGTGAC 1624

RESULT 5
AX135707
LOCUS AX135707
DEFINITION Sequence 26 from Patent WO0132878.
ACCESSION AX135707
VERSION AX135707.1 GI:14271957
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 2013)
AUTHORS Rowe, P.
TITLES Polypeptide hormone phosphatonin
JOURNAL Patent: WO 0132878-A 26 10-MAY-2001;

```

UNIVERSITY COLLEGE LONDON (GB)
 FEATURES Location/Qualifiers
 source . . . 1..2013
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 740 a 367 c 459 g 447 t
 ORIGIN

Alignment Scores:
 Pred. NO.: 2, 21e-122 Length: 2013
 Score: 2276.00 Matches: 429
 Percent Similarity: 100.00% Conservatave: 1
 Best Local Similarity: 99.77% Mismatches: 0
 Query Match: 99.87% Indels: 0
 Gaps: 0

US-09-700-696b-2 (1-430) x AX135707 (1-2013)

QY 1 ValAsnLysGluTyrSerIleSerAsnLysGluAsnThrHisAsnGlyLeuArgMetSer 20
 ::
 Db 359 CTGAATAAGAAATATAGTATCATCAAGAAAGAAATACTACAAATGGCTGAGAGATCTCA 418
 QY 21 IleTyrProLysSerThrGlyAsnLysGlyPheGluAspGlyAspAlaIleSerLys 40
 ::
 Db 419 ATTTATCTTAAGTCAACTGGGAATTAAGCGTTTGAGAGATGAGATGATGCTATCAGCAA 478
 QY 41 LeuHisAspGlnGluTyrGlyAlaAlaLeuIleArgAsnAsnMetGlnHisIleMet 60
 ::
 Db 479 CTACATGACCAAGAAAGAAATATGGCGAGCTCTCATCAAGAAATTAACATGCACAAATATATG 538
 QY 61 GlyProValThrAlaIleLysLeuLysGluLysGluAsnLysGluAsnThrProArgAsn 80
 ::
 Db 539 GGGCCAGAGAGATGCGATTAACTCTGGGGAGAGAAACAAAGAAAGAACACACCTAGAGAT 598
 QY 81 ValLeuAsnIleIleProAlaSerMetAsnTyrAlaLysAlaHisSerLysAspLysLys 100
 ::
 Db 599 GTTCTAAACATATATCCCAAGCAAGTATGTAATATGCTTAAGCAACACTCAAGATTAAGAAAG 658
 QY 101 LysProGlnArgAspSerGlnAlaGlnLysSerProValLysSerLysSerThrHisArg 120
 ::
 Db 659 AAGCTCAAGAGATTTCCCAAGCCCAAGAAAGTCCAGTAAGAAAGCAAGCAAGCAAGCTATGCT 718
 QY 121 IleGlnHisAsnIleAspTyrLeuLysHisLysSerLysValLysLysIleProSerAsp 140
 ::
 Db 719 ATTCACACACAAATGACTACTCTAAACATCTCTCAAAAAGTCAAAAAAATCCCCAGTGAT 778
 QY 141 PheGluGlySerGlyTyrThrAspLeuGlnArgGlyAspAsnAspIleSerProPhe 160
 ::
 Db 779 TTTGAAGGACGCGTTATACAGATCTTCAAGAGAGGAGCAAGTATATATCTCTCTTTC 838
 QY 161 SerGlyAspGlyGlnProPheLysAspIleProGlyLysGlyGluAlaThrGlyProAsp 180
 ::
 Db 839 ACTGGGAGGCGCACTTTTAAGGACATTTCTGGTAAAGGAAAGCTACTGCTGCTGAC 898
 QY 181 LeuGluGlyLysAspIleGlnThrGlyPheAlaGlyProSerGluAlaGlnSerThrHis 200
 ::
 Db 899 CTGGAAGGCAAGATATTAACAAAGCGTTTGCAGGCCCAAGTGAAGCTGAGACTACTCAT 958
 QY 201 LeuAspThrLysLysProGlyTyrAsnGluIleProGluArgGluGluAsnGlyLysAsn 220
 ::
 Db 959 CTGGACACAAAAAGCAGGTTATATATGAGATCCCGAAGAGAAAGAAATAGTGGAAAT 1018
 QY 221 ThrIleGlyThrArgAspGluThrAlaLysGluAlaAspAlaValAspAlaSerLeuVal 240
 ::
 Db 1019 ACCATTGGAACTAGGATGAACCTGGAAAGAGGACAGATGCTGATGTCAGCCCTTGA 1078
 QY 241 GluGlySerAsnAspIleMetGlySerThrAsnPheLysGluLeuProGlyArgGlyGly 260
 ::
 Db 1079 GAGGGAGCAAGCATATCATGGGTAGTACCAATTTTAAGAGAGCTCCCTGAAAGAGAGA 1138
 QY 261 AsnArgValAspAlaGlySerGlnAsnAlaHisGlnGlyLysValGluPheHisTyrPro 280
 ::
 Db 1139 AACAGATGATGCTGGCAGCCAAAGTCTACCAAGGAGAGGTGAGTTTCATTACCT 1198

QY 281 ProAlaProSerLysGluLysArgLysGluGlySerSerAspAlaIleGlnSerThrAsn 300
 ::
 Db 1199 CCTGCACCTCTCAAAAGAGAAAGAAAGAGAGCAGTAGATGATGAGTGAAGTACCAAC 1258
 QY 301 TyrAsnGluIleProLysAsnGlyLysGlySerThrArgLysGlyValAspHisSerAsn 320
 ::
 Db 1259 TATATGAATTCCTTAAAAATGGCAAGCAGTACCAGAAAGGGGTGATGATCTTAAAT 1318
 QY 321 ArgAsnGluAlaThrLeuAsnGluLysGlnArgPheProSerLysGlyLysSerGlnGly 340
 ::
 Db 1319 AGGAACCAAGCAACCTTAATGAAAAACAAAGATTCTTCATAGTAAGGGCAAAAGTCAGGGC 1378
 QY 341 LeuProIleProSerArgLysLeuAspAsnGluIleLysAsnGluMetAspSerPheAsn 360
 ::
 Db 1379 CTGCCCATTCCTCTCGCTGCTGTGATATGAATCAAAAAGCAATGATTCCTTAAAT 1438
 QY 361 GlyProSerHisGluAsnIleIleThrHisGlyArgLysTyrHisTyrValProHisArg 380
 ::
 Db 1439 GGGCCAGATCATGAGATATTAATTAACACATGCGAGAAATATCATTTATGTACCCACACA 1498
 QY 381 GlnAsnAsnSerThrArgAsnLysGlyMetProGlnGlyLysGlySerTrpGlyArgGln 400
 ::
 Db 1499 CAAATATATTCTACAGGAATTAAGGATATGCCCAAGGAAAGCGCTCGGGGTAGACAA 1558
 QY 401 ProHisSerAsnArgArgPheSerSerArgArgArgAspAspSerSerGluSerSerAsp 420
 ::
 Db 1559 CCCCATTCCAACAGAGAGGTATGTTCCCGTAGAAGGAGATGACAGTATGATCATCTGAC 1618
 QY 421 SerGlySerSerSerGluSerAspGlyAsp 430
 ::
 Db 1619 AGTGGCAGTTCAAGTGAAGCGATGCTGAC 1648

RESULT 6
 AC093768 187624 bp DNA linear PRI 01-MAR-2002
 LOCUS
 DEFINITION Homo sapiens BAC clone RP11-113613 from 4, complete sequence.
 AC093768 AC093768 GI:16973767
 VERSION
 AC093768.3 GI:16973767
 KEYWORDS
 HTG.
 SOURCE
 Homo sapiens.
 ORGANISM
 Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE
 1 (bases 1 to 187624)
 Harkins, C., Kozlowski, A. and Doeber, A.
 The sequence of Homo sapiens BAC clone RP11-113613
 Unpublished (2001)
 JOURNAL
 3 (bases 1 to 187624)
 Toward a complete human genome sequence
 Genome Res. 8 (11), 1097-1108 (1998)
 JOURNAL
 MEDLINE
 99063792
 PUBMED
 9847074
 2 (bases 1 to 187624)
 Harkins, C., Kozlowski, A. and Doeber, A.
 The sequence of Homo sapiens BAC clone RP11-113613
 Unpublished (2001)
 JOURNAL
 3 (bases 1 to 187624)
 Toward a complete human genome sequence
 Genome Res. 8 (11), 1097-1108 (1998)
 JOURNAL
 MEDLINE
 99063792
 PUBMED
 9847074
 3 (bases 1 to 187624)
 Harkins, C., Kozlowski, A. and Doeber, A.
 The sequence of Homo sapiens BAC clone RP11-113613
 Unpublished (2001)
 JOURNAL
 3 (bases 1 to 187624)
 Toward a complete human genome sequence
 Genome Res. 8 (11), 1097-1108 (1998)
 JOURNAL
 MEDLINE
 99063792
 PUBMED
 9847074
 4 (bases 1 to 187624)
 Harkins, C., Kozlowski, A. and Doeber, A.
 The sequence of Homo sapiens BAC clone RP11-113613
 Unpublished (2001)
 JOURNAL
 3 (bases 1 to 187624)
 Toward a complete human genome sequence
 Genome Res. 8 (11), 1097-1108 (1998)
 JOURNAL
 MEDLINE
 99063792
 PUBMED
 9847074
 5 (bases 1 to 187624)
 Harkins, C., Kozlowski, A. and Doeber, A.
 The sequence of Homo sapiens BAC clone RP11-113613
 Unpublished (2001)
 JOURNAL
 3 (bases 1 to 187624)
 Toward a complete human genome sequence
 Genome Res. 8 (11), 1097-1108 (1998)
 JOURNAL
 MEDLINE
 99063792
 PUBMED
 9847074
 6 (bases 1 to 187624)
 Harkins, C., Kozlowski, A. and Doeber, A.
 The sequence of Homo sapiens BAC clone RP11-113613
 Unpublished (2001)
 JOURNAL
 3 (bases 1 to 187624)
 Toward a complete human genome sequence
 Genome Res. 8 (11), 1097-1108 (1998)
 JOURNAL
 MEDLINE
 99063792
 PUBMED
 9847074

AUTHORS
 TITLE
 JOURNAL
 COMMENT

Waterston, R.
 Direct Submission
 Submitted (01-MAR-2002) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 On Nov 17, 2001 this sequence version replaced g1:16418232.

----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc
 Contact: sapiens@wustl.wustl.edu/gsc
 Summary Statistics
 Center project name: H_NH0113G13
 Drafting Center: WIBR

NOTICE: This sequence may not represent the entire insert of this
 clone. It may be shorter because we only sequence overlapping
 clone sections once, or longer because we provide a small overlap
 between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate
 chemistry, or covered by high quality data (i.e., phred quality >
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by sequence
 from more than one subclone; and the assembly was confirmed by
 restriction digest.

MAPPING INFORMATION:
 Mapping information for this clone was provided by Dr. John D.
 McPherson, Department of Genetics, Washington University, St. Louis
 MO. For additional information about the map position of this
 sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
 The RPI1-11 human BAC library was made from the blood of one male
 donor, as described by Osogawa, K., Moon, P. Y., Zhao, B., Fritgen, E.,
 Taitano, M., Catanesi, J. J., and de Jong, P. J. (1996) An improved
 approach for construction of bacterial artificial chromosome
 libraries. Genomics 51:1-8. The clone may be obtained either from
 Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
 and co-workers at http://www.chori.org
 VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:
 The clone sequenced to the right is AC087106. Actual start of this
 clone is at base position 1 of RPI1-113G13; actual end is at base
 position 187624 of RPI1-113G13.

Sequence derived from one plasmid subclone, base position 69516 to
 69532.

The sequence of AC021959 has been incorporated into AC093768.

FEATURES
 source
 Location/Qualifiers

1..187624
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="4"
 /map="4"
 /clone="RPI1-113G13"
 /clone_11b="RPI1-11"
 6069..6195
 repeat_region
 /rpt_family="L1"
 6288..6308
 repeat_region
 /rpt_family="AT_rich"
 6422..6530
 repeat_region
 /rpt_family="L1"
 6559..6582
 repeat_region
 /rpt_family="AT_rich"
 6708..6744
 repeat_region
 /rpt_family="AT_rich"
 6732..7028
 repeat_region
 /rpt_family="Alu"

repeat_region
 8033..8053
 /rpt_family="(TTTAA)n"
 repeat_region
 8041..8459
 /rpt_family="L1"
 repeat_region
 8904..8942
 /rpt_family="MIR"
 repeat_region
 8943..9214
 /rpt_family="MALR"
 repeat_region
 9526..9591
 /rpt_family="(TA)n"
 repeat_region
 10055..10530
 /rpt_family="L1"
 repeat_region
 11326..11417
 /rpt_family="L2"
 repeat_region
 11399..11445
 /rpt_family="(TGA)n"
 repeat_region
 15645..15865
 /rpt_family="(TGA)n"
 repeat_region
 15933..16227
 /rpt_family="L2"
 misc_feature
 complement(16160..16161)
 /note="similar to Homo sapiens EST A1745652 (NID:g5113940)
 tr24c10.x1"
 repeat_region
 16226..16251
 /rpt_family="Alu"
 misc_feature
 16337..16436
 /note="similar to Homo sapiens EST AA873157 (NID:g2969279)
 oh7q11.s1"
 misc_feature
 16340..16915
 /note="similar to Homo sapiens EST AV702108
 (NID:g10718458)"
 misc_feature
 16340..16899
 /note="similar to Sus scrofa EST B1181092 (NID:g14655501)"
 16340..16447
 /note="similar to Rattus norvegicus EST BF391035
 (NID:g11375882)"
 misc_feature
 16340..16436
 /note="similar to Homo sapiens EST AA651669 (NID:g2583321)
 ns89e04.s1"
 misc_feature
 16340..16436
 /note="similar to Homo sapiens EST AA865538 (NID:g3000646)
 ns22d02.s1"
 misc_feature
 16340..16436
 /note="similar to Homo sapiens EST A1088133 (NID:g3427109)
 c285907.x1"
 misc_feature
 16340..16436
 /note="similar to Homo sapiens EST AL533242
 (NID:g12796735)"
 misc_feature
 16340..16436
 /note="similar to Homo sapiens EST BE742618
 (NID:g10156597)"
 misc_feature
 16340..16436
 /note="similar to Homo sapiens EST B1094057
 (NID:g14512387)"
 misc_feature
 16342..16436
 /note="similar to Homo sapiens EST AA398444 (NID:g2051756)
 zt74g07.s1"
 misc_feature
 16355..16380
 /note="similar to Homo sapiens EST W00667 (NID:g1272102)
 za72b10.r1"
 misc_feature
 16388..16436
 /note="similar to Homo sapiens EST BE621830
 (NID:g9892770)"
 misc_feature
 16403..17080
 /note="similar to Homo sapiens EST B1084966
 (NID:g14503296)"
 misc_feature
 16405..16436
 /note="similar to Homo sapiens EST B1522324
 (NID:g15347116)"
 misc_feature
 16406..16436
 /note="similar to Homo sapiens EST A1745652 (NID:g5113940)
 tr24c10.x1"
 misc_feature
 16458..17080

```

/Note="similar to Homo sapiens EST AL533242
(NID:912796735)"
misc_feature
16458..17080
/Note="similar to Homo sapiens EST BE621830
(NID:98922770)"
misc_feature
16458..17080
/Note="similar to Homo sapiens EST BE742618
(NID:910156597)"
misc_feature
16458..17077
/Note="similar to Homo sapiens EST B1522324
(NID:915347116)"
misc_feature
16458..16873
/Note="similar to Homo sapiens EST AA398444
(NID:92051756)
2574907.s1"
16458..16868
/Note="similar to Homo sapiens EST B1094057
misc_feature
Alignment Scores:
Pred. No.: 3.68e-120 Length: 187624
Score: 2276.00 > Matches: 429
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.77% Mismatches: 0
Query Match: 99.87% Indels: 0
Gaps: 0
DB: 9
US-09-700-696b-2 (1-430) x AC093768 (1-187624)
OY 1 ValAsnlysglytyrserlleSerAsnlysgluasnThrHisAsnlygluArgmetSer 20
:::|||||
Db 75853 CTGATAAAGATATAGATATAGTACCAAGAAATCTCATCTGAGATGCTGAGATGCA 75912
OY 21 lleyrProlyserthrlyAsnlysglyphegluaspglyaspapalilSerlys 40
|||||
Db 75913 ATTHTCTAGTCAACGCGGAATGAAGGTTTGAGATGAGATGATGATGACCAA 75972
OY 41 leuHisaspngluyluglytyrGlyAlaAlaLeuileArgAsnAsnMetGlnHislleMet 60
|||||
Db 75973 CTACATGACCAAGAAATATGCGCAGCTCTCATCAGAAATTAACATCAATATATG 76032
OY 61 GlyProvalThrAlaileLysLeuLeuylgluLysAsnlysgluasnThrProArgsn 80
|||||
Db 76033 GGGCCAGTACCTCGCATTAATCTCTGGGGGAGAAACCAAGAAACACACTAGAT 76092
OY 81 valLeuasnilleleProAlaSerMetAsnThrAlaLysAlaHisSerLysAspLys 100
|||||
Db 76093 GTTCTAAACATATCCACAGATATGATTTGCTAAAGCACACTCGAAGATATAAAG 76152
OY 101 LysProGlnArgAspSerGlnAlaGlnLysSerProValLysSerLysSerThrHisArg 120
|||||
Db 76153 AGCCCAAGAGATCCCAAGCCGAGAAAGCCAGTAAAAAGCAAAAGCACCCATCGT 76212
OY 121 lleGlnHisAsnllleAspThrLeuLysHisleuSerLysValLysLyslleProSerAsp 140
|||||
Db 76213 ATTCAACACACATGACTACTTAAACATCTCTCAAAAGTCAAAAAATCCCACTGAT 76272
OY 141 pheGlnlyserglytyrThrAspLeuGlnLysArgLysAspAsnAspLysSerPrope 160
|||||
Db 76273 TTGTAAGCCAGCGTTTATACAGATCTTCAAGAGAGAGGAGCAATATATCTCTTTC 76332
OY 161 serGlyaspGlyGlnProphelyAspLysPheProGlyLysGlyLysAlaThrGlyProasp 180
|||||
Db 76333 AGTGGGAGCGCCAACTTTAAGGACATCTCGTAAAGAGAGAGCTACTGCTCTGAC 76392
OY 181 leuGlnlyLysAspLysGlnThrGlyPheAlaGlyProSerGlyLysAlaGlnSerThrHis 200
|||||
Db 76393 CTAAAGGCAAGATATTCACAAAGGCTTGACAGGCCCAAGTGAAGCTGAGAGTACTCAT 76452
OY 201 leuAspThrLysLysProGlyTyraSngluileProGluArgLysGlnGlyLysAsn 220
|||||
Db 76453 CTGACACAAAAAGCCAGGTTATATGAGATCCAGAGAGAGAAAGAGGAGTGAAT 76512
OY 221 ThrllleGlyThrArgAspGluThrAlaLysGlnLysAlaLysAlaLysValSerLeuVal 240
|||||

```

```

Db 76513 ACCTTGGAACTAGGATGAAGAACTGCAAGAGAGCAAGATCTGTGATGACCTTGA 76572
OY 241 GluGlySerAsnAspLysMetGlySerThrAsnPhelyGluLeuProGlyArgGly 260
|||||
Db 76573 GAGGCGAGCAACGATATCAGGTAGTACCAATTTTAAGAGAGCTCCCTGGAGAGAAAGA 76632
OY 261 AsnArgValAspAlaGlySerGlnAsnAlaHisGlnlyLysValGluPheHisTyPro 280
|||||
Db 76633 AACACAGTGCATGCTGGAGAGCAAAATGCTCCACAGAGAGAGGTTGAGTTTCATACCT 76692
OY 281 ProAlaProSerLysGluLysArgLysGluGlySerSerAspAlaAlaGlnSerThrAsn 300
|||||
Db 76693 CCGTGCACCCCTCAAAAAGAAAAAGAAAGAGAGGAGAGGAGAGGAGTGAAGTCCCTTAT 76752
OY 301 TyrAsnGluileProLysAsnGlyLysGlySerThrArgLysGlyValAspHisSerAsn 320
|||||
Db 76753 TTTAATGAAATTCCTAAATATGGCAAGGAGTACCAAGAAAGGGGTGATGATCTTAT 76812
OY 321 ArgAsnGlnAlaThrLeuAsnGluLysGlnArgPheProSerLysGlyLysSerGlnly 340
|||||
Db 76813 AGGACCAAGCAACCTTAATATGAAAAACAAAGGTTCTAGTAAGGCAAAAGTCAGGGC 76872
OY 341 leuProLysProSerArgGlyLysAspAsnGluileLysAsnGluMetAspSerPheAsn 360
|||||
Db 76873 CTGCCATTCCTCTCTGCTGCTGCTGATATGAAATCAAAAGCAATGATTCCTTAT 76932
OY 361 GlyProSerHisGluAsnilleleThrHisGlyArgLysTyHisTyValProHisArg 380
|||||
Db 76933 GGGCCAGATCATGAAATTAATTAACATGCGCAGAAATATCATATGTACCCACAGA 76992
OY 381 GlnAsnAsnSerThrArgAsnLysGlyMetProGlnlyLysGlySerThrPylArgGln 400
|||||
Db 76993 CAAATTAATTCACAGGATTAAGGTATGCCACAAAGGAAAGGCTCCGGGGTGAACA 77052
OY 401 ProHisSerAsnArgArgPheSerSerArgArgArgAspAspSerSerLysSerAsp 420
|||||
Db 77053 CCCCATTCCAAGAGAGGTGTAGTCCCGTAAAGAGGATGACAGTATGATCATCTGAC 77112
OY 421 SerGlySerSerGlySerAspGlyAsp 430
|||||
Db 77113 AGTGGAGTTCAAGTGAAGAGATGCTGAC 77142
RESULT 7
AB046056 2095 bp mRNA linear PRI 09-AUG-2000
LOCUS AB046056
DEFINITION Macaca fascicularis mRNA for matrix extracellular
phosphoglycoprotein(MEPE gene), complete cds.
ACCESSION AB046056.1 GI:9280167
VERSION AB046056.1
KEYWORDS f1s (Full Insert sequence).
SOURCE Macaca fascicularis adult cDNA to mRNA, clone_1lb:macaque brain
cDNA library Occe clone:Occe-12450.
ORGANISM Macaca fascicularis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopithecinae; Macaca.
REFERENCE 1 (sites)
Osada, N., Hida, M., Kusuda, J., Tanuma, R., Iseki, K., Hirai, M.,
Terao, K., Suuki, Y., Sugano, S., and Hashimoto, K.
Isolation of full-length cDNA clones from macaque brain cDNA
libraries
Unpublished
2 (bases 1 to 2095)
Hashimoto, K., Osada, N., Hida, M., Kusuda, J. and Sugano, S.
Direct Submission
Submitted (14-JUL-2000) Katsuyuki Hashimoto, National Institute of
Infectious Diseases, Division of Genetic Resources; 23-1, Toyama
1-chome, Shinjuku-ku, Tokyo 162-8640, Japan
(E-mail: khashi@nih.go.jp, URL: http://www.nih.go.jp/yoken/genebank/,
Tel: 81-3-5285-1111 (ex. 2120), Fax: 81-3-5285-1181)
URL: http://www.nih.go.jp/yoken/genebank/
Lib Name: macaque brain cDNA library Occe
Lab host: TOP10
COMMENT

```


Vector: pME185-FL3 (Acc.No. AB009864)
 R. Site1: DraIII (CACTGTCG)
 R. Site2: DraIII (CAGCATGTCG)
 Description: 1st strand cDNA was primed with an oligo(dT) primer using specific 5' and 3' primers and amplified by PCR. The PCR product was digested with SfiI and size selection was performed to exclude fragments <1.5kb. The SfiI-digested PCR product was cloned into distinct DraIII sites of pME185-FL3. XhoI sites just outside the DraIII sites can be used to isolate the cDNA insert. Libraries were constructed by Sugano et al. (University of Tokyo, Institute of Medical Science). Custom primer used for sequencing
 (5' end primer (CTTCTGCTCTAAAGCTGCG);
 3' end primer (CGACCTGACGTCGACACAA)).
 Location/Qualifiers
 1. 2095
 /organism="Macaca fascicularis"
 /db_xref="taxon:9541"
 /clone="QcCE-12430"
 /clone_11b="macaque brain cDNA library QcCE"
 /dev_stage="adult"
 69. 1736
 /codon_start=1
 /product="matrix extracellular phosphoglycoprotein"
 /protein_18="BAB01638.1"
 /db_xref="GI:9280168"
 /translation="MRFVCVGLFLSYWMAAPFQPTQTEKQSCVEORITTKGHE
 KHGVFEVCYVNSPKKQNDVQOEKKNQDIGHHIGKRRYQELSKENITVQERKD
 LSLESENNSSKSONFTNRRLNKYSIKENHNGIKRMTYRKSTGNKQFAG
 DDATSKLHDOEYGAALIRNMQHIKMPVIAIKLGEENKSKRPYNLKIIPASMANA
 KAHSKDKKPPORDSOVQKVPYKSKSTHRTQNIIDYPHLSKVKIPDFEGSGTDLQ
 ERGNDSPEFGDQPRKDIPEGKEATSDLEKDIQTFAGSEASTNLDTEPEY
 NEIPERKENGNTIGTDETAKEADAVDSIEGNDIMGSTEINKELPGEENVDVG
 GQNAHQKVEFHYPAPSKERKEKSSDASTSTYNIENPKNDGSAKRGVDDSRNA
 ILHKORPPSKSGOGLPIPSRGLNDEIKEMDSLNPSNENPHSKRYIVYVHRON
 PTRNKGPHGSGKSGRQPSNRRLSSRRSSSDSSGSSSDGQ"

FEATURES

source

CDS

BASE COUNT 783 a 381 c 471 g 460 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 7.28e-110 Length: 2095
 Score: 2058.50 Matches: 391
 Percent Similarity: 93.95% Conservative: 13
 Best Local Similarity: 90.93% Mismatches: 25
 Query Match: 90.32% Indels: 1
 Gaps: 1
 US-09-700-696b-2 (1-430) x AB046056 (1-2095)

OY 1 ValAsnLysGluTyrSerIleSerAsnLysGluAsnThrHisAsnGlyLeuArgMetSer 20
 Db 447 CTGATTAAGAATAATAGTATCAGTACAAAGAGAAATATTCACAAATGCGCTAGAGATGCA 506
 OY 21 IleTyrProLysSerThrGlyAsnLysGlyPheGluAspGlyAspAspAlaIleSerLys 40
 Db 507 ATTATCTAGTCACTCGGAAATAAACAGTTTCGCGATGGAGTGTATGCTATTCAGCAA 566
 OY 41 LeuHisAspGlnGluTyrGlyAlaLeuIleArgAsnAsnMetGlnHisIleMet 60
 Db 567 CTACATGACCAAGAAATATGCGCGAGCTCTCATGAAATATACATGCAACATATATG 626
 OY 61 GlyProValThrAlaIleLysLeuGlyGluGluAsnLysGluAsnThrProArgAsn 80
 Db 627 GGGCCAGTCACTGCAATTAATCTCTGGGGAAGAAACAAACAGACCAACCTAAGAAC 686
 OY 81 ValLeuAsnIleLeuProAlaSerMetAsnTyrAlaLysAlaHisSerLysAspLys 100
 Db 687 GTCTTAACAATAATTCAGCAGATATGATCTCTTAAGCAGACCTGGAAGATATAAG 746
 OY 101 LysProGlnArgAspSerGlnAlaGlyLysSerProValLysSerLysSerThrHisArg 120
 Db 747 AAGCTCAAGAGATTCCTCAAGATCCAGAAAGTTCCAGTAAAGCAACACCATCTCGT 806

OY 121 IleGlnHisAsnIleAspTyrLeuLysHisLeuSerLysValLysLysIleProSerAsp 140
 Db 807 ACTCAACACAAACATGATGATACCCAAACATCATCAAAAGTCMAAAATATCCACAGTAT 866
 OY 141 PheGluLysSerGlyTyrThrAspLeuGlnLysArgGlyAspAsnAspIleSerProPhe 160
 Db 867 TTTGAGGACGAGGTTACACAGATCTTCAAGAGAGAGGCAATGATATCTCTCTTC 926
 OY 161 SerGlyAspGlyGlnProPheLysAspIleProGlyLysGlyGlyAlaThrGlyProAsp 180
 Db 927 AGTGGGAGGCGCAACCTTTTAAGGACATTCCTGTAAAGAGAGAGACTACTGTTCTGAC 986
 OY 181 LeuGluGlyLysAspIleGlnThrGlyPheAlaGlyProSerGlyAlaGlySerThrHis 200
 Db 987 CTAGAGGCAAGATATTCAAACAGGCTTTCAGCCCAAGTCAAGCTCGAGATCTAAT 1046
 OY 201 LeuAspThrLysLysProGlyTyrAsnGluIleProGluArgGluGluAsnGlyLysAsn 220
 Db 1047 CTGACACAAAGAGCCAGGTTATATGATGATCCCAAGAAAGAAATATGGCGAAT 1106
 OY 221 ThrIleGlyThrArgAspGlyThrAlaLysGluAlaAspAlaValAspValSerLeuVal 240
 Db 1107 ACCATTGGAACTGGGATGAAACTGCGAAAGAGCAGATGCTGTATGTCAGTCTTGA 1166
 OY 241 GluLysSerAsnAspIleMetGlySerThrAsnPhelysGluLeuProGlyArgGly 260
 Db 1167 GAGGCAACACAGATATCATGATGATGATCAATTTTAAGCACTCCCTCGAAGAGAGCA 1226
 OY 261 AsnArgValAspAlaGlySerGlnAsnAlaHisGlnLysValGluPheHisTyrPro 280
 Db 1227 AACAGATGATGATGTGGCGGCCCAAAATGCTCACCAAGGAGAGTTGATTTTACCT 1286
 OY 281 ProAlaProSerLysGlyLysArgGlyGlyLysSerSerAspAlaIleGluSerThrAsn 300
 Db 1287 CTGTCACCTTCACAAAGAGAAAGAGAGAGCAGATGATGATCACTCACTGAATACCAAT 1346
 OY 301 TyrAsnGluIleProLysAsnGlyLysGlySerThrArgLysGlyAlaAspHisSerAsn 320
 Db 1347 TATATGAATATTCCTTAAGAAAGCAGAAAGGCGAGAGGAGGTATGATGATCTTAAT 1406
 OY 321 ArgAsnGlnAlaThrLeuAsnGluLysGlnArgPheProSerLysGlyLysSerGlnGly 340
 Db 1407 AGAACCACCAACATCTTACATGAAACAAAGTTCTCTAGTAAAGGCAAAATGTCAGGCC 1466
 OY 341 LeuProIleProSerArgGlyLeuAspAsnGluIleLysAsnGluLeuAspSerPheAsn 360
 Db 1467 CTGCCCATTCCTCTGCTGCTTGTATGATGATGAAATCAAAACGGAATGATTCCTTAAT 1526
 OY 361 GlyProSerHisGluAsnIleIleThrHisGlyArgLysTyrHisTyrValProHisArg 380
 Db 1527 GGGCCAGTATAGAT---ATACCAACACAGCAGAAATATCATTAATATACCCATAGA 1583
 OY 381 GlnAsnAsnSerThrArgAsnLysGlyMetProGlnGlyLysGlySerProLysArgGln 400
 Db 1584 CAAATAATCTCTACAGGATTAAGGTATGCCACATGAGAAAGGCTCCCTGGGGTACACA 1643
 OY 401 ProHisSerAsnArgArgPheSerSerArgArgArgAspAspSerSerGluSerAsp 420
 Db 1644 CCAATATCCACAGGAGGCTTATGTTCCCTTAAGAGAGAGACAGTATGATCATCTGAC 1703
 OY 421 SerGlySerSerSerGluSerAspGlyAsp 430
 Db 1704 AGTGGCAGTTCAAGTACAGCATGCTGAC 1733
 RESULT 8
 AB060891
 LOCUS 2259 bp mRNA linear PRI 13-JUN-2001
 DEFINITION Macaca fascicularis brain cDNA clone:qlra-13588, full insert
 sequence.
 ACCESSION AB060891
 VERSION AB060891.1 GI:13874559
 KEYWORDS Macaca fascicularis adult male temporal lobe right cDNA to mRNA,
 SOURCE

ORGANISM

clone.lib:macaque brain cDNA library QtrA clone:QtrA-13588.
 Macaca fascicularis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 Cercopithecinae; Macaca.

REFERENCE

1 (sites)
 Osada,N., Hida,M., Kusuda,J., Tanuma,R., Iseki,K., Hirai,M.,
 Terao,K., Suzuki,Y., Sugano,S. and Hashimoto,K.

TITLE

Isolation of full-length cDNA clones from macaque brain cDNA
 libraries

JOURNAL

Unpublished

REFERENCE

2 (bases 1 to 2259)

AUTHORS

Hashimoto,K., Osada,N., Hida,M., Kusuda,J. and Sugano,S.

TITLE

Direct Submission

JOURNAL

Submitted (27-APR-2001) Katsuyuki Hashimoto, National Institute of
 Infectious Diseases, Division of Genetic Resources; 23-1, Toyama
 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan
 (E-mail:hashim@nih.go.jp, url:http://www.nih.go.jp/yoken/genebank/,
 Tel:81-3-5285-1111(ex.2120), Fax:81-3-5285-1181)

COMMENT

Lab host: TOP10
 Vector: pME18S-FL3 (Acc.No. AB009864)
 R. Site1: DraIII (CACATGTG)
 R. Site2: DraIII (CACATGTG)
 Description: 1st strand cDNA was primed with an oligo(dt) primer
 [ATGTCGCTTTTCTTTTCTTTT], double-stranded cDNA was synthesized
 using specific 5' and 3' primers and amplified by PCR. The PCR
 product was digested with SfiI and size selection was performed to
 exclude fragments <1.5kb. The SfiI-digested PCR product was cloned
 into distinct DraIII sites of pME18S-FL3. XhoI sites just outside
 the DraIII sites can be used to isolate the cDNA insert. Libraries
 were constructed by oligo-capping method
 (Sugano et al., Institute of Medical Science, University of
 Tokyo).

Custom primer used for sequencing
 (5' end primer [CTTCTGCTCTTAAAGCTGCG];
 3' end primer [CGACCTGACGTCGACACA]).

FEATURES

Location/Qualifiers

1..2259

/organism="Macaca fascicularis"

/db_xref="taxon:9541"

/clone="QtrA-13588"

/sex="male"

/tissue-type="temporal lobe right"

/clone.lib="macaque brain cDNA library QtrA"

/dev_stage="adult"

381..1883

/codon_start=1

/product="hypothetical protein"

/protein_id="BAB46894.1"

/db_xref="GI:13874560"

/translation="MSPGKKNQTDVKEKKNKNDIGLHLGKRRYQELSKENIVQER
 KKDLSISEASENNGSKSONYFNINROINKESISNKENIHGLMSIYPKSTGNKOF
 ADGDALISKLHDEYGAALFENNNOHIMPTATKLIGEEKKOSKPNVLRKAKF
 NTAHAKSKDKKKPQSDQVQKVPYKSKSHRQNHIDVPHISKYKTIKPSDESGSYT
 DLQERGDNDMSPSGDQGFQDIPKGEETGPDLDKDIQTFAFPAESAEINDDTKE
 PGYNEIPEKEENGNTIGTDETAKEAVNVSLVGGNNDIGSTNEFLPERBENRY
 DVGCGNAHOGKVEFYHPAPSKERKREGSDATESITNNEIPKNDKGSARKVDVDSNR
 NOAILHEKORFEPKSKSGSLPIPSRGLDNEITKENDSLNGPENEIPISRKHYVPHR
 QNNPFRNKGMPHGKSGMWGROPYSNRLSSRRRDESESSDSSSESDD"

CDS

BASE COUNT

836 a 424 c 505 g 494 t

ORIGIN

Alignment Scores:
 Pred. No.: 7, 92e-110 Length: 2259
 Score: 2058.50 Matches: 391
 Percent Similarity: 93.95% Conservative: 13
 Best Local Similarity: 90.93% Mismatches: 25
 Query Match: 90.32% Indels: 1
 Gaps: 1

US-09-700-696b-2 (1-430) x AB060891 (1-2259)

QY 1 ValasnlsglufyrSerlleSerAsnlysgluasnthHisasngilyLeuAqgmetSer 20

```

Db 594 CTGAATRAAGAAATATGATACATCAACAAAGAGAAATATTCACCAATGGCTGAGGATGCA 653
QY 21 lletyrProlySerThrGlyAsnlyGlyPheGluAspGlyAspAspAlaIleSerlyS 40
Db 654 ATTATCTTAAGCAACTGAGGATTAACACTTTGGCGGATGGAATGATGCTATACACAA 713
QY 41 LeuHisaspGlnGluGlyGlyAlaLeuIleArgAsnmetGlnHisIleMet 60
Db 714 CTGATGACCAAGAGAAATATGCGCGACGCTCTCATGAAATATGCAATATTAAG 773
QY 61 GlyProvalThrAlaIleLeuLeuGlyGluGluAsnlyGluAsnThrProAlaAsn 80
Db 774 GGGCACTGATGCAATTAACCTCGGAGAAACCAACAGCAACCACTAGAAC 833
QY 81 ValLeuAsnIleIleProAlaSerMetAsnTyAlaIleAsnHisSerlyAspIlyS 100
Db 834 GTTCTTAACAAATATTCGAGAGATGAACTATGCTTAACACACACTGAAAGATAAAG 893
QY 101 LysProGlnArgAspSerGlnAlaGlnLysSerProValLysSerlySerThrHisArg 120
Db 894 AAGCTCAAGAGATTCCTCCAAAGTCCAGATAAAGCAAAACCAACCACTCGT 953
QY 121 lIeGlnHisAsnIleAspTyTLeuLysHisLeuSerlyValLysIlyProSerAsp 140
Db 954 ACTCAACACAAACATTTGACTACCCAAACATCTATCAAAAGCAAAATCCCACTGAT 1013
QY 141 PheGluGlySerGlyTyTThrAspLeuGlnGluArgGlyAspAsnAspIleSerProPhe 160
Db 1014 TTGAGGACGCGGTTCACAGATCTTCAAGAGAGGGGCAATGATGTCCTTC 1073
QY 161 SerGlyAspGlyGlnProPheLysAspIleProGlyLysGlyAlaIleThrGlyProAsp 180
Db 1074 AGTGGGACGCGCCAACTTTTAAAGACATTCCTATTAAGAGAGAGTACTGCGCTGAC 1133
QY 181 LeuGluGlyAspIleGlnThrGlyPheAlaGlyProSerGlyAlaGlySerThrHis 200
Db 1134 CTAGAGGCAAAAGATATTCACAGAGGTTTCAGAGCCCAATGAAAGTGAGATCAAT 1193
QY 201 LeuAspThrLysLysProGlyTyTAsnGluIleProGluArgGluAsnGlyGlyAsn 220
Db 1194 CTGACACAAAGAGACCGAGTATATGATCCACAGAAAGAGAAAGAAAGGCGGAAT 1253
QY 221 ThrIleGlyThrArgAspGluThrAlaLysGluAlaAspAlaValAspValSerLeuVal 240
Db 1254 ACCATTGGAACCTGGGAGAACTGCGAAGAGGACAGTGTCTTAAGTCAAGTCTTCTA 1313
QY 241 GluGlySerAsnAspIleMetGlySerThrAsnPheLysGluLeuProGlyArgGluGly 260
Db 1314 GAGGCGCAACAGATATCATGCTGATCCCAATTTAAGCAACTCCGGAAGAGAGGA 1373
QY 261 AsnArgValAspAlaGlySerGlnAsnAlaHisGlnGlyLysValGluPheHisTyPro 280
Db 1374 AACACAGTGGATGTGGCGGCGCAAAACGCTCCCAAGGAGAGGTGAGTTCATTAACCT 1433
QY 281 ProAlaProSerLysGluLysArgLysGluGlySerSerAspAlaAlaGlySerThrAsn 300
Db 1434 CCTGCACCTCTCAAGACAGAAAGAAAGAGAGCAATAGATGCAACGAAAGTACCAAT 1493
QY 301 TyTAsnGluIleProLysAsnGlyLysGlySerThrArgLysGlyValAspHisSerAsn 320
Db 1494 TATATGAATTCCTTAAGACAGCAAGGAGCAAGGAGTGTAGATACCTAAT 1553
QY 321 ArgAsnGlnAlaThrLeuAsnGluLysGlnArgPheProSerLysGlyLysSerGlnGly 340
Db 1554 AGCAACCAAGCAATCTTACATGAAAGAAAGAGGTTCTAGTAAGGCCAAAGTACAGGC 1613
QY 341 LeuProIleProSerArgGlyLeuAspAsnGluIleLysAsnGluMetAspSerPheAsn 360
Db 1614 CTGCCATTCCTCTCTGCTGCTGATGATGAATGAATGAATGAATGAATGATCCCTAAT 1673
QY 361 GlyProSerHisGluAsnIleIleThrHisGlyArgLysTyHisTyAlaProHisArg 380

```

Db 1674 GGGCCAGTATGAGAAAT--ATACACACAGCAGAAAATATCATTTATGATACCCCATAGA 1730
 Qy 381 GlnAsnAsnSerThrArgAsnLysGlyMetProGlyGlyGlySerProGlyArgGln 400
 Db 1731 CAAATTAATCTACACAGGATTAAGGATATCCCATGGGAAAGGCTCGGGGTAGACAA 1790
 Qy 401 ProHisSerAsnArgArgPheSerSerArgArgArgAspAspSerSerGluSerSerAsp 420
 Db 1791 CCTATTCCAAACAGGAGCGTTAGTCCCTAGAAAGGACAGACAGTAGTAGTCACTGTGAC 1850
 Qy 421 SerGlySerSerGluSerArgAspGlyAsp 430
 Db 1851 AGTGCAGTTCACAGTACGACGATGCTGAC 1880
 RESULT 9
 LOCUS AB050259
 DEFINITION Macaca fascicularis brain cDNA, clone: Qnpa-21045.
 ACCESSION AB050259
 VERSION AB050259.1 GI:10998236
 KEYWORDS fls (full insert sequence).
 SOURCE Macaca fascicularis adult male brain parietal lobe cDNA to mRNA, clone, lib: macaque brain cDNA library Qnpa clone: Qnpa-21045.
 ORGANISM Macaca fascicularis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea; Macaca.
 REFERENCE
 AUTHORS Osada, N., Hida, M., Kusuda, J., Tanuma, R., Iseki, K., Hirai, M., Terao, K., Suzuki, Y., Sugano, S. and Hashimoto, K.
 TITLE Isolation of full-length cDNA clones from macaque brain cDNA libraries
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 2140)
 AUTHORS Hashimoto, K., Osada, N., Hida, M., Kusuda, J. and Sugano, S.
 TITLE Direct Submision
 JOURNAL Submitted (20-OCT-2000) Katsuyuki Hashimoto, National Institute of Infectious Diseases, Division of Genetic Resources; 23-1, Toyama 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan
 (E-mail: khashimoto@nih.go.jp, URL: http://www.nih.go.jp/yoken/genebank/, Tel: 81-3-5285-1111 (ex. 2120), Fax: 81-3-5285-1181)
 COMMENT
 Lab host: TOP10
 Vector: pME18S-FL3 (Acc. No. AB009864)
 R. Site1: DraIII (CACTGTGTC)
 R. Site2: DraIII (CACCATGTC)
 Description: 1st strand cDNA was primed with an oligo(dT) primer (ATGTCCTTTTCTTTTCTTTTCTTTT); double-stranded cDNA was synthesized using specific 5' and 3' primers and amplified by PCR. The PCR product was digested with SfiI and size selection was performed to exclude fragments <1.5kb. The SfiI-digested PCR product was cloned into distinct DraIII sites of pME18S-FL3. XhoI sites just outside were constructed by Sugano et al. (University of Tokyo, Institute of Medical Science). Custom primer used for sequencing (5' end primer [CTTGTGCTCTTAAGCTGCG]; 3' end primer [CGACCTGACGTCGACGAC]).
 Location/Qualifiers
 1. 2140
 /organism="Macaca fascicularis"
 /db_xref="taxon:9541"
 /clone="Qnpa-21045"
 /sex="male"
 /tissue_type="brain parietal lobe"
 /clone_lib="macaque brain cDNA library Qnpa"
 /dev_stage="adult"
 110..1777
 /codon_start=1
 /product="MEPE protein"
 /protein_id="BAB17010.1"
 /db_xref="GI:10998237"
 /translation="MRVECVGLFLSYMAAPRPPQREKTKQSCVEEDRRTTYKHHE KHGVYFKCVVMSPKKNOTDQKDEKKNQDGLHGLRRRTQESKSNENTVOERKRD LSLSEGENNGSSKSONFTFNKORLNKEYSISNKENIHNLMSIYPRSTGNKOPADG

Alignment Scores:
 Pred. NO.: 1 27e-109 Length: 2140
 Score: 2054.50 Matches: 390
 Percent Similarity: 93.95% Conservative: 14
 Best Local Similarity: 90.70% Mismatches: 25
 Query Match: 90.15% Indels: 1
 DB: 9 Gaps: 1
 US-09-700-696b-2 (1-430) x AB050259 (1-2140)
 Qy 1 ValAsnLysGlySerThrIleSerAsnLysGluAsnThrHisAsnGlyLeuArgMetSer 20
 Db 488 CTGAATAAAGATATAGTATCAGTAACAAGAGAAATATCAATGCGCTGAGAGATCTCA 547
 Qy 21 IleThrProLysSerThrGlyAsnLysGlyPheGluAspGlyAspAlaIleSerLys 40
 Db 548 ATTATCTTAAGTCACTGAGTAATTAACACTTGGCGGATGAGATGCTCTATCAGCGAA 607
 Qy 41 LeuHisAspGlnGluGlyArgAlaAlaLeuIleArgAsnAsnMetGlnHisIleMet 60
 Db 608 CTACATGACCAAGAAGATATGGCGGACGCTCTATCAGAAATACATTCACATTAATAG 667
 Qy 61 GlyProValThrAlaIleLysLeuGlyGluGluAsnLysGluAsnThrProArgAsn 80
 Db 668 GGGCAGTCACTCAATTAATTAACCTCGGGGAGAAACAGACCAACCTAGAAC 727
 Qy 81 ValLeuAsnIleIleProAlaSerMetValArgAlaLysAlaHisSerLysAspLys 100
 Db 728 GTTCTTAACCAAAATTCAGCAACTATGCTTAAGCACACACGTAAGGATTAAG 787
 Qy 101 LysProGlnArgAspSerGlnAlaGlnLysSerProValLysSerLysSerThrHisArg 120
 Db 788 AACCTCAAGAGATGCCAAGTCCAGAAAGTTCCAGTAAAGCAAAAGCACCATGCT 847
 Qy 121 IleGlnHisAsnIleAspThrLeuLysHisLeuSerLysValLysIleProSerAsp 140
 Db 848 ACTCAACACACACTTACTACCCCAAAACATCTATCAAAAGTCAAAATAATCCCAAGTAT 907
 Qy 141 PheGluGlySerGlyThrAspLeuGlnGluArgGlyAspAsnAspIleSerProPhe 160
 Db 908 TTGAAGGACAGCGGTATACACAGATCTTCAAGAGAGGACAAATGATATGCTCTTTC 967
 Qy 161 SerGlyAspGlyGlnProPheLysAspIleProGlyLysGlyGluAlaThrGlyProAsp 180
 Db 968 AGGGGAGAGGCAACCTTTAAGGACATCTCGTAAAGGAGCAAGCTACGCTGTGAC 1027
 Qy 181 LeuGluGlyLysAspIleGlnThrGlyPheAlaGlyProSerGluAlaGluSerThrHis 200
 Db 1028 CTAGAAGGCAAAAGATATTAACACAGGTTTGCAGGCCCAAGGAGCTGAGAGTACTAT 1087
 Qy 201 LeuAspThrLysLysProGlyLysArgGlnLysGluIleProGluArgGluGluAsnGlyLys 220
 Db 1088 CTTGACACAAAAGAGCGGTATATGATGATCCAGAAAGAAAGAAATGGCGGAAT 1147
 Qy 221 ThrIleGlyThrArgAspGluThrAlaLysGluAlaAspAlaValAspValSerLeuVal 240
 Db 1148 ACCATTGGAACTGGGAGTAAACTCGAAAGGCGAGATGCTTGAGATGAGCTTCTTA 1207
 Qy 241 GluGlySerAsnAspIleMetGlySerThrAsnPheLysGluLeuProGlyArgGluGly 260
 Db 1208 GAGGCAACAACAGATATCATGCGTAGTACCAATTTTAAGAAACCTCCCGAAGAGAAAGA 1267
 Qy 261 AsnArgValAspAlaGlySerGlnAsnAlaHisGlnGlyLysValGluProHisIleTyPro 280


```

Oy 106 SerGlnIaGlnIySerProValIySer-----IySerThrHisArgIleGln 122
Db 419 CTGCTAACCCAGAGCAGCCCGGTCACAAAGCAACACACCAACACACCCGACACCCGA 478
Oy 123 HisAnIleAspIyLeuIyHisLeuSerIyValIyLysIleProSerAspPheGlu 142
Db 479 CGGACCACTACCTACCTACACATCTCCACAGATCAAGAGAGCTCCGAGACTTGAA 538
Oy 143 GlySerGlyTyrThrAspLeuGlnIyArgIyAspAsnAspIleSerProPheSerGly 162
Db 539 GGCAGTGGCTCCCGCATCTTCTACTGAGGAGGATATATGATGTCCTCCCTTCAGTGA 598
Oy 163 AspGlyGlnProPheIyAspIleProGlyIyGlyIyGlnIyValIyThrGlyProAspLeu 181
Db 599 GATGGCAACATTTATATGACATCTCTGCAAGAGAGGCTGGGTCTGCTCTGCA--- 655
Oy 182 GluIyIyAspIleGlnIyThrGlyPheAlaGlyProSerGlnIyGlnIySerThrHisLeu 201
Db 656 -----ACGTCACACTAGTCGCCCCCTCAGGCTCAGCAAGAGTGAAGTTATTGACCCA 709
Oy 202 AspThrIyLysIyProGlyTyrAsnGlnIleProGluIyArgGluIyGlnIyValIyThr 221
Db 710 CATATGATGAGCTAGGCTCTTATGATGATCCCGGAGAGAGAGACATGTCGAGTGC 769
Oy 222 IleGlyThrArgAspIyThrAlaIyGlnIyAlaAspAlaValIyAspValIySerLeuValGlu 241
Db 770 TATGACACAGAGACAAACCTGACAGGGGCGAGCTGTCAGTGGGAGCTTGCGG 829
Oy 242 GlySerAspIleMetIySerThrAspPheIyGlnIyLeuProGlyIyArgGluIyAsn 261
Db 830 GGCAGCAATGAATTCACAGAGCAGCACCACCAATTCAGGGAATCTCCCGGAAAGAGAGAAC 889
Oy 262 ArgValAspAlaGlySerGlnAsnAlaHisGlnIyIyValIyGluPheHisIyTyrProPro 281
Db 890 AGAATTATGCGCGAGCAAAATGCTCATCAAGGAAAGTGAATTCATATCCACAA 949
Oy 282 AlaProSerIyGlnIyIyArgIyGlnIyIySerSerAspAlaAlaGlnIySerThrAsnIy 301
Db 950 GTGCGCTCGAGCAAAAGCTAAAGGGGGGCTGAGCAGTCAAGGAGACT---GGTTAC 1006
Oy 302 AsnGlnIleProIyAsnGlnIyIyGlySerThrArgIyGlnIyAlaIyAspHisSerAsnArg 321
Db 1007 AACGAAATCCCAAGAGCAGCAAGAGTCTAGCAGCAAGATCAGCAAGAGTCCAAAGG 1066
Oy 322 AsnGlnAlaThrLeuAsnGlnIyIyGlnIyArgPheProSerIyGlnIyIySerGlnIyLeu 341
Db 1067 AACCAATTACCTTGACTGCAAGCAGCAAGATTCACAGTAAAGGCAAAAGCCAGGCGCT 1126
Oy 342 ProIleProSerArgIyLeuAspAsnGlnIleIyAsnGlnIyMetAspSerPheAsnIy 361
Db 1127 GCTGTGCTCTCAGCTCTTATGATGAGTTAAAGTAA----- 1168
Oy 362 ProSerHisGlnAsnIleIleThrHisGlyArgIyIyTyrHisIyValIyProHisArgGln 381
Db 1169 -----GAAAC-----CATATGTTCTCATGAGCA 1195
Oy 382 AsnAsnSerThrArgAsnIyIyGlyMetProGlnIyIyGlySerIyTyr---GlyArgGln 400
Db 1196 AATATCTTACCAAGCAATTAAGGATGTCAAGGAGGAGGCTCTGCTGCTGAGAGAGA 1255
Oy 401 ProHisSerAsnArgArgPheSerSerArgArgAspAspSerSerSerSerSerSerSer 420
Db 1256 CCCATTTCACAGCGCGCTGACAGCCGCGCAAGA---GACAGCAGCAGATGCTCATCC 1312
Oy 421 SerGlySerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 430
Db 1313 AGTGGAGTCTAGTGAAGTCAATGTCAC 1342
RESULT 12 AF298661 1682 bp mRNA linear ROD 27-JUL-2002
LOCUS AF298661 Mus musculus osteoregulin mRNA, complete cds.

```

```

ACCESSION AF298661 GI:21998850
VERSION AF298661.1
KEYWORDS house mouse, Mus musculus
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 1682)
AUTHORS Cowen, L.C., Petersen, D.N., Vail, A.L., Stock, J., Tralcevic, G.T., Simmons, H.A., Childsey-Frink, K.L., Ke, H., McNeish, J., and Brown, T.A.
TITLE Targeted disruption of the osteoregulin gene results in increased bone density
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1682)
AUTHORS Brown, T.A.
TITLE Direct Submission
JOURNAL Submitted (23-AUG-2000) Cardiovascular & Metabolic Disease, Pfizer Global Research & Development, Eastern Point Road, Groton, CT 06340, USA
FEATURES
SOURCE location/Qualifiers
1..1682
/db_organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="5"
/map="5E3-E5"
67..1392
/note="extracellular matrix protein; expressed in osteoblasts and osteocytes of bone; similar to rat osteoblast/osteocyte factor 45"
CDS
1/codon_start=1
/product="osteoregulin"
/protein_id="AA067687.1"
/db_xref="GI:21998851"
/translation="MTPEGLMKQAVSVGLLPSMTAAAPNEDRSSGNGDSIHKD
LAASVYSDPVDGTEGCGALDHPDQDHYGALLRNTQPKSLVTAELRREGNQ
EKROVSTVTPADNDVNAKIKDIDKNOESYLTOSSPYSKTKHTRTSTHYLT
HLRQIKTPSDLESBSGPDLLVGDNDNVPFSGDGHFMHIIPEGAGSGQESTSRP
LSSGSKAEVIDPHNSGLSNEITREBHGSAATADKAAQAGASGSLVGSNEIT
GSTNFRLEPKREGRINAGSONAHQGVKVERHTFOVASREKVKGVNHAAGACVNEIPK
SSKSSSKDAEESKGNLTLTASQRFPGKRSQGPALPSHLSNEKVSERNHVFPHQ
NNTLPNKGMQSQRCSMPSRPNSHRRASFTQRDSSESSSSSSSHSGD"
BASE COUNT 510 a 401 c 432 g 339 t
ORIGIN
Alignment Scores:
Pred. No.: 3,33e-48 Length: 1682
Score: 983.50 Matches: 217
Percent Similarity: 62.56% Conservative: 52
Best Local Similarity: 50.47% Mismatches: 134
Query Match: 43.15% Indels: 27
DB: Gaps: 8
US-09-700-696b-2 (1-430) x AF298661 (1-1682)
Oy 6 SerIleSerAsnIyGlnAsnThrHisAsnGlnIyLeuArgMetSerIleTyrProIySer 25
Db 166 AGCTGGGCAATCAGACAGCATTCACAGAGACTCTGTGATCTGATCC 225
Oy 26 ThrGlyAsnIyGlnIyPheGlnIyAspAlaIleSerIyLeuHisAspGlnIy 45
Db 226 AGCGTAGATGAAGGACAGAGAGATGGGAGGTCTCTCTTCAACCCGCTGGCCAGAGAC 285
Oy 46 GluTyrGlyAlaIleAlaIleuIleArgAsnAsnMetGlnHisIleMetGlyProValIyAla 65
Db 286 AGCTATGTCCTCCCTCCACAGAAATATACAGCAACCTCTAAAGAGTACTAGTGGG 345
Oy 66 IleIyLeuIyGlnIyGlnIyAsnIyIyGlnIyThrProArgAsnValIyLeuAsnIleIle 85
Db 346 GCGGACTAGAGGAGAGAGAAACACAGAGAGACTCAAGCTGTTTAAGCCGTAAT 405
Oy 86 ProAlaSerMetAsnIyIyAlaIyAlaHisSerIyAspIyGlnIyIySerProIyIyArgAsp 105
Db 406 CCAGCAGATGTCATGATGCTTAAAGTCTCTTAAAGACATATAAGATCAAGAGATTAT 465

```

```

Oy 106 serGlnAlaGlnLysSerProValLysSer-----LysSerThrHisArgIleGln 122
Db 466 CCGCAATACCCAGACAGCCCGGTCTCAAAACCAACACACACACACCGCCAGCCGCA 525
Oy 123 HisAnIleAspTyrLeuLysHisLeuSerLysValLysLysIleProSerAspPheGlu 142
Db 536 CGGACACTCATCTGCTGACATCTCCACAGATCAAGAAAGACTCCACAGTACCTTAA 585
Oy 143 GlysGlyTyrThrAspGlnGlnLysArgLysAspAsnAspIleSerProPheSerGly 162
Db 586 GGCATGCTGCTCCCGACATCTTCTACTGAGGAGATTAATGATGTCCTTCAGTGA 645
Oy 163 AspGlyLysProPheLysAspIleProGlyLysGlyGluAla---ThrGlyProAspLeu 181
Db 646 GATGGCGAAACATTTTATGACATTCCTGCAAGAGAGTGGTGGTCTGCTGCTGAA--- 702
Oy 182 GluGlyLysAspIleGlnThrGlyPheAlaLysProSerGluAlaGluSerThrHisLeu 201
Db 703 -----ACCTCACTAGTCCGCCCTCTCAGCGCTCCAGCAAAAGCTGAAGTATGACCCA 756
Oy 202 AspThrLysLysProGlyTyrAsnGluIleProGluArgGluGlnLysAsnThr 221
Db 757 CATATGACTGGACTAGCTTAATGAGATCCCGGGAGACAGACATGGGCGAGTGC 816
Oy 222 IleGlyThrArgAspGluThrAlaLysGluAlaAspAlaValAspValSerLeuValGlu 241
Db 817 TATCAACACAGACAGCAAAAGCTGACAGGGGCGAGCTCTCAGATGGGACCTTGTCGG 876
Oy 242 GlysSerAsnAspIleMetGlySerThrAsnPheLysGluLeuProGlyArgGluGln 261
Db 877 GGCAGCATGAATACACAGCAGCAGCAGCAATTCAGGAGATCCCGGAAAAAGAGAAC 936
Oy 262 ArgValAspAlaGlySerGlnAsnAlaHisGlnLysValGluPheHisTyrProPro 281
Db 937 AGAATTTATCTCCGCGACCAAAATGCTCATCAAGGAAAGTACAGATTCTCATCCACA 996
Oy 282 AlaProSerLysLysArgLysGluGlySerSerAspAlaAlaGluSerThrAsnTyr 301
Db 997 GTGGCCCTCGAGAGAAAGGTAAGGGGGGGGCTGACAGCATCAGGAGAGCT---GGTTAC 1053
Oy 302 AsnGluIleProLysAsnGlyLysGlySerThrArgLysGlyValAlaAspHisSerAsnArg 321
Db 1054 AACGAATTCCTCCAGAGCAGCAAAAGTACCTAGCAAAAGATCAGAGATCCCAAGG 1113
Oy 322 AsnGlnAlaThrLeuAsnGlnLysGlnArgPheProSerLysGlyLysSerGlnGlyLeu 341
Db 1114 AACCAATTACCTTGAAGTACAGCCCAAAAGTTTTCAGGTAAAGGCAAAAGCCAGGCCCT 1173
Oy 342 ProIleProSerArgLysLeuAspAsnGluIleLysAsnGluMetAspSerPheAsnGly 361
Db 1174 GCTCTGCTCCCTCAGAGTCTTATGATGAGTTAAAGTGA----- 1215
Oy 362 ProSerHisGluAsnIleIleThrHisGlyArgLysTyrHisTyrValProHisArgGln 381
Db 1216 -----GAAAC-----CATTATGTGTTTCATGAGACA 1242
Oy 382 AsnAsnSerThrArgAsnLysGlyMetProGlnGlyLysGlySerTyr---GlyArgGln 400
Db 1243 AATAATCTTACACCGAATAAAGGATGTCACAGCGAGAGCTCTGCGCTTCGAGAGA 1302
Oy 401 ProHisSerAsnArgArgPheSerSerArgArgAspAspSerSerSerSerSerAsp 420
Db 1303 CCCAATTCCACAGCGCGCTGACACCGCCCAAGA---GACAGACAGCAGTCTGATCC 1359
Oy 421 SerGlySerSerSerGluSerAspLysAsp 430
Db 1360 ACTGGAGTTCTTAGTAGAGTATGATGTAGC 1389

```

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS

TITLE
JOURNAL
REFERENCE
AUTHORS
JOURNAL

COMMENT

***, 62 unordered pieces.
AC129695
AC129695.1 GI:22038416
HTG; HTGS_PHASE1.
Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 169603)
Muzny, D.M., Adams, C., Adio-Oduola, B., All-ouman, F.R., Allen, C.,
Alstbrooks, S.L., Amaratunga, H.C., Are, J.R., Ayala, M., Banks, T.,
Barber, J., Benton, J., Blinze, K., Blankenburg, K., Bonaldi, D.,
Bouch, J., Bowie, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P.,
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
Carroll, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
Chen, G., Chen, R., Chen, Z., Chowdhury, I., Christopoulos, C.,
Cleveland, C.D., Cox, C., Goye, M.D., Dathorne, S.R., David, R.,
Dayala, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
Hernandez, O., Hodgson, A., Hognes, M., Hollway, C., Hollins, B.,
Homs, F., Howard, S., Huber, J., Hulik, S., Hume, J., Jackson, L.E.,
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudh, S.,
Karlssohn, E., Kelly, S., Khan, U., King, L., Korvan, J., Kovar, C.,
Kratovic, J., Kuresh, A., Landy, N., Leal, B., Lewis, L.C., Lewis, L.,
Li, J., Li, Z., Lichtarge, O., Liu, C., Liu, J., Liu, M., Louised, H.,
Lorado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mel, G., Metzger, M.,
Miner, G., Miner, Z., Mitchell, T., Mohabath, K., Morgan, M., Morris, S.,
Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokoko, S., Ogun, K., Okunolu, G.,
Orgunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peterson, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
Rivers, M., Rojas, A., Rojuben, I., Rolfe, M., Ruiz, S., Sawyer, G.,
Scherer, S., Scott, G., Shen, H., Shoochari, N., Sisson, I.,
Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Swalek, A., Taber, P., Tamara, A., Tamara, K., Tang, H.,
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, R., Thomas, S.,
Umani, K., Vasquez, L., Vera, Y., Villalob, C., Vinson, R., Wang, Q.,
Williams, G., Williamson, A., Warren, R., Washington, C., Wellington, S.,
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 169603)
Submitted (01-AUG-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GMA
Center clone name: CH230-11B16
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 12061 bases at least Q40
Consensus quality: 126182 bases at least Q30
Consensus quality: 131508 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 62 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1 1069: contig of 1069 bp in length
* 1070 1169: gap of unknown length
* 1170 2314: contig of 1145 bp in length
* 2315 2414: gap of unknown length
* 2415 3550: contig of 1136 bp in length
* 3551 3650: gap of unknown length
* 3651 4963: contig of 1313 bp in length
* 4964 5063: gap of unknown length
* 5064 6271: contig of 1208 bp in length
* 6272 6371: gap of unknown length
* 6372 7857: contig of 1486 bp in length
* 7858 7957: gap of unknown length
* 7958 9006: contig of 1049 bp in length
* 9007 9106: gap of unknown length
* 9107 10720: contig of 1614 bp in length
* 10721 10820: gap of unknown length
* 10821 11853: contig of 1033 bp in length
* 11854 11953: gap of unknown length
* 11954 13115: contig of 1162 bp in length
* 13116 13215: gap of unknown length
* 13216 14346: contig of 1131 bp in length
* 14347 14446: gap of unknown length
* 14447 15977: contig of 1531 bp in length
* 15978 16077: gap of unknown length
* 16078 17689: contig of 1612 bp in length
* 17690 17789: gap of unknown length
* 17790 18887: contig of 1198 bp in length
* 18888 19087: gap of unknown length
* 19088 21446: contig of 2359 bp in length
* 21447 21546: gap of unknown length
* 21547 23482: contig of 1936 bp in length
* 23483 25723: gap of unknown length
* 25724 25823: gap of unknown length
* 25824 27149: contig of 1326 bp in length
* 27150 28663: contig of 1414 bp in length
* 28664 28763: gap of unknown length
* 28764 30632: contig of 1869 bp in length
* 30633 30732: gap of unknown length
* 30733 32080: contig of 1348 bp in length
* 32081 32180: gap of unknown length
* 32181 34021: contig of 1841 bp in length
* 34022 34121: gap of unknown length
* 34122 36060: contig of 1939 bp in length
* 36061 36160: gap of unknown length
* 36161 37317: contig of 1157 bp in length
* 37318 37417: gap of unknown length
* 37418 39691: contig of 2274 bp in length
* 39692 39791: gap of unknown length
* 39792 41839: contig of 2048 bp in length
* 41840 41939: gap of unknown length
* 41940 43817: contig of 1878 bp in length
* 43818 43917: gap of unknown length
* 43918 46079: contig of 2162 bp in length
* 46080 46179: gap of unknown length
* 46180 48668: contig of 2489 bp in length
* 48669 48768: gap of unknown length
* 48769 50979: contig of 2211 bp in length
* 50980 51079: gap of unknown length
* 51080 53320: contig of 2241 bp in length
* 53321 53420: gap of unknown length

```

```

* 53421 55272: contig of 1852 bp in length
* 55273 55372: gap of unknown length
* 55373 57327: contig of 1955 bp in length
* 57328 57427: gap of unknown length
* 57428 59613: contig of 2186 bp in length
* 59614 59713: gap of unknown length
* 59714 61816: contig of 2103 bp in length
* 61817 61916: gap of unknown length
* 61917 64312: contig of 2396 bp in length
* 64313 64412: gap of unknown length
* 64413 66627: contig of 2215 bp in length
* 66628 66727: gap of unknown length
* 66729 69791: contig of 3064 bp in length
* 69892 74354: contig of 4463 bp in length
* 74355 74454: gap of unknown length
* 74455 77688: contig of 3414 bp in length
* 77689 77968: gap of unknown length
* 77969 80955: contig of 2987 bp in length
* 80956 81055: gap of unknown length
* 81056 84342: contig of 3287 bp in length
* 84343 84442: gap of unknown length
* 84443 86371: contig of 1929 bp in length
* 86372 86471: gap of unknown length
* 86472 87673: contig of 1202 bp in length
* 87674 87773: gap of unknown length
* 87774 91770: contig of 3997 bp in length
* 91771 91871: gap of unknown length
* 91872 94538: contig of 2668 bp in length
* 94539 94638: gap of unknown length
* 94639 97030: contig of 2392 bp in length
* 97031 97130: gap of unknown length
* 97131 103898: contig of 6768 bp in length
* 103899 107372: gap of unknown length
* 107373 107472: contig of 3374 bp in length
* 107473 110167: gap of unknown length
* 110168 110267: contig of 2695 bp in length
* 110268 111267: gap of unknown length
* 111268 114227: contig of 3860 bp in length
* 114228 114227: gap of unknown length
* 114229 118441: contig of 4214 bp in length
* 118442 118541: gap of unknown length
* 118542 121332: contig of 2791 bp in length
* 121333 121432: gap of unknown length
* 121433 127282: contig of 5850 bp in length
* 127283 127382: gap of unknown length
* 127383 132945: contig of 5563 bp in length
* 132946 133045: gap of unknown length
* 133046 139139: contig of 6094 bp in length
* 139140 139240: gap of unknown length
* 139241 142147: contig of 2908 bp in length
* 142148 142247: gap of unknown length

```

Alignment Scores:

```

Pred. No.: 2.81e-44 Length: 169603
Score: 954.50 Matches: 214
Percent Similarity: 60.66% Conservative: 45
Best Local Similarity: 50.12% Mismatches: 143
Query Match: 41.88% Indels: 25
DB: Query Match: 2 Gaps: 8

```

US-09-700-696b-2 (1-430) x AC129695 (1-169603)

```

QY 9 AaLysGLuAsnThrHisAsnGlyLeuArgMetSerIleTyrProLysSerThrGlyAsn 28
   |||||  |||  |||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1072261 AACCAAGGACATGTCAC-----TTAGCATGCGTAAACGACGCCGCTGGGT 102311

QY 29 LysGLyPheGluAspGlyAspAspAlaIleSerLysIleHisAspGluGluIuTyrGly 48
   |||||  |||  |||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 102312 AAGGAACAGAGGCTGGAGATGCTCCCTTACCTGCTGACCAAGACAGGAGGCT 102371

QY 49 AlaAlaLeuIleArgGlnAsnMetGlnHisIleMetGlyProValThrAlaIleLysIleu 68
   |||  |||||  |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```



```

Db 102372 GCCACCTCTCAGAAATATCATCTCAGCTGTAAAGAGTCTGTGACGGGAGCTGAGTA 102431
Qy 69 LeuGlyGluGlnAsnLysGluAsnThrProArgAsnValLeuAsnIleIleProAlaSer 88
Db 102432 CAGAGCGCAGAAACAAAGAAAGAAACCTCAGAGTGTCTTAAGCCGTAATTCACACAGAT 102491
Qy 89 MetAsnIlyrAlaLysAlaIleSerLysAspLysLysLysProGlnArgAspSerGlnAla 108
Db 102492 GTCCACAAATGCTAAAGCACTACTCCGAGATACAGAGAACACAGAGGAGTCTACTACTC 102551
Qy 109 GlnLysSerProValLysSerLysSerThrHisArgIleGlnHisAsnIleAspTyrLeu 128
Db 102552 CAGAACGCCAGCAGACAGAACACACACCTCTCGGGCCGAGACAGACGACCTACCTA 102611
Qy 129 LysHisLeuSerLysValLysLysIleProSerAspPheGluGlnSerGlyTyrThrAsp 148
Db 102612 ACACATCTCCCAAAATCAGAAATCTTCAGAGACTTCAGAGACAGTGTCTCCACAGAC 102671
Qy 149 LeuGlnGluArgLysAspAsnAspIleSerProPheSerGlyAspLysGlnProPheLys 168
Db 102672 CTTCCTAGTGGGGGAGATATGATGTCCTCTTCAGTGGAGATGAGACACATTTTATG 102731
Qy 169 AspIleProGlyLysGlyGluAlaThrGlyProAspLeuGlu-----GlyLysAsp 185
Db 102732 CACACTCCGACAGAGAGAGTGTCTGATCTGATCTGAAAGCTCAGCTGTGACCT 102791
Qy 186 IleGlnThrGlyPheAlaGlyProSerGluAlaGluSerThrHisLeuAspThrLysLys 205
Db 102792 GTC-----TCAGCTCCAGCATGTGAGATGTTCAGAGATGTTCACCCACACAGATGGA 102839
Qy 206 ProGlyTyrAsnGluIleProGluIleArgGluGlnGlyAsnThrIleGlyThrArg 225
Db 102840 CTGGCTCTAATGATGATCCAGAGAGAGAGTGCATAGCGGCTGCTATGCAACAGAGA 102899
Qy 226 AspGluThrAlaLysGluAlaAspAlaValAspAlaSerLeuValGluGlySerAsnAsp 245
Db 102900 GGAAAAACGCCAGAGGGGAGGCTCCGGGATGTGACGCTTGTGAGAGGACCAATGAA 102959
Qy 246 IleMetGlySerThrAsnPheLysLysLeuProGlyArgGluGlyAsnArgValAspAla 265
Db 102960 ATCAGAGGCGCATGCCAAATTTAGGGAGCTCCCTGGAAAAAGAGAAACAGAGTCGATCC 103019
Qy 266 GlySerGlnAsnAlaHisGlnGlyLysValGluPheHisTyrProProAlaProSerLys 285
Db 103020 AGCAGCCAAATGCTCATCAGAAAGTAATTCCTACCTCCCAAGCCGCTCAAAA 103079
Qy 286 GlnLysArgGlyGlySerSerAspAlaAlaGluSerThrAsnTyrAsnGluIlePro 305
Db 103080 GAGAAAGTAAAGGGGGCAGCAGGAGCAGACAGGAAAGCCGGTTTCAATGAATATCCCC 103139
Qy 306 LysAsnGlyLysGlySerThrArgLysGlyValAspHisSerAsnArgAsnGlnAlaThr 325
Db 103140 AAGAGCAGCAAGGCGGCGCAGCTACGACAGAGATCCGGAAAGAAATCTAAAGGAAACCAAGTAAC 103199
Qy 326 LeuAsnGluLysGlnArgPheProSerLysGlyLysSerGlnGlyLeuProIleProSer 345
Db 103200 TTGACTGAAAGCCAAAGTTCCAGGCAAGGCAAGGCCAGCT-----TCT 103247
Qy 346 ArgGlyLeuAspAsnGluIleLysAsnGluMetAspSerPheAsnGlyProSerHisGlu 365
Db 103248 CACAGCTTGTGTAATGAGGTAAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAG 103307
Qy 366 AsnIle---IleThrHisGlyArgLysTyrHisTyrValProHisArgGlnAsnAsnSer 384
Db 103308 GCGATTTCATATACACACAGAGAGAACAGCCAC-----CCT 103343
Qy 385 ThrArgAsnLysGlyMetProGlnGlyLysGlySerTrp---GlyArgGlnProHisSer 403
Db 103344 ACACGGATATGGGGGATGTCACAGCGAGGCTCTGCGCTCGAGAGAGACCCATCC 103403
Qy 404 AsnArgArgPheSerSerArgArgAspAspSerSerGlnSerSerAspSerGlySer 423
Db 103404 CACCGCGCGCTAAGCAGACCGCCCAAGA-----GACAGTAGTAGTATCATCATCCAGTGGAGCT 103460

```

```

Qy 424 SerSerGluSerAspGlyAsp 430
Db 103461 TCTAGCGAGAGCAGCTGTGATC 103481

RESULT 14
AF530559
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
1 (bases 1 to 1651)
Wang, X., Hu, B. and Wang, Y.
Rattus norvegicus cDNA sequence expressed in B4 cell line (possible
subtype of osteoregulin)
unpublished
2 (bases 1 to 1651)
Wang, X., Hu, B. and Wang, Y.
Direct Submission
Submitted (15-JUL-2002) Radiation Oncology, Thomas Jefferson
University, 1020 Sansom St. Thompson Bldg. B-13, Philadelphia, PA
19107, USA

FEATURES
source
location/Qualifiers
1..1651
/organism="Rattus norvegicus"
/strain="F344"
/db_xref="taxon:10116"
/cell_line="B4"
14..1345
/codon_start=1
/product="osteoregulin-like protein"
/protein_id="AA094404.1"
/db_xref="GI:22212816"
/translation="MQLFLSQLKTEHREPKTGKCHKLNEDSGSGNGNTHLA
SVKPEPMGKGTGGADAPLRLDDNRQCATLRLNTPYKSLYGTGYQSDRNEKK
POSLVSVITDYHNTDYSSEDFENORQDLNLSGSGSKHTPRARSTHYLTHLPQIR
KILSDPEDASPDLLVRGNDYPPSSGSDQHMHNPDRGAASDPESSAGHPVSGSS
NLEIYDPRHNGISGVNIPRBEHIGATATKGTAKQAGASADVSLVEGSENETGSTK
RELPGKEGRNVDASSQNAHQKVERHYDQAPSKYKVGSGSEHETGACYNETPPSSKG
GASKDAESKGVLTLSORPRPGKGSLSLGNENYKSEEDSSNSLRGIAMHR
RTSHPTRNKMSORGSWASRPHPHRRVSTRQRODSSESSSSSESSSDP"

BASE COUNT 529 a 384 c 433 g 305 t
ORIGIN

Alignment Scores:
Pred. No.: 1.26e-45 Length: 1651
Score: 938.50 Matches: 211
Percent Similarity: 59.95% Conservative: 45
Best Local Similarity: 49.41% Mismatches: 146
Query Match: 41.18% Indels: 25
DB: 10 Gaps: 8

US-09-700-696B-2 (1-430) x AF530559 (1-1651)
Qy 9 AsnLysGluAsnThrHisAsnGlyLeuArgMetSerIleTyrProLysSerThrClyAsn 28
Db 122 AACCAAGCAGACATCCAC-----TTAGCATGTGTGAAGCCTGAGCCCATGTGGCT 172
Qy 29 LysGlyPheGlnAspGlyAspAspAlaIleSerLysLeuHisAspGlnGluTyrGly 48
Db 173 AAAGCAACAGAGGGTGGCGAGATGCTCCCTTCACCTGCTTGACACAGAACAGCAGAGGT 232
Qy 49 AlaAlaLeuIleArgAsnAsnMetGlnHisIleMetGlyProValIleThrAlaIleLysLeu 68
Db 233 GCCACCCCTCAGAAATATATCATCAGCTCTTAAGAGACTGTGGTACGGGAGCTGAAGTA 292
Qy 69 LeuGlyGluGlnAsnLysGluAsnThrProArgAsnValLeuAsnIleIleProAlaSer 88

```



```

Db 293 CAGAGCGACGAGAAACAGAAAGAAACCTCAGAGTGTCTTACGATTAATCCACAGAT 352
OY 89 MetantTyrAlaLysAlaHisSerLysAspLysLysProGlnArgAspSerGlnAla 108
Db 353 GTCCACAACTACTACACACTACTACAGAGTACAGAAACACAGAGGATCTACTACTC 412
OY 109 GlnLysSerProValLysSerLysSerThrHisArgIleGlnHisAsnLysPyrLeu 128
Db 413 CAGAACAGCCCGAGCAAAAGAACACACCTCGGGCCCGAGCAAGCAGCACTACTACTA 472
OY 129 LysHisLysSerLysValLysLysLysLysProSerAspPheGluLysSerGlyTyrThrAsp 148
Db 473 ACACATCTCCCAATTCAGAAAGATTTCAGACTTCGAGACAGTCTCTCCACAC 532
OY 149 LeuGlnGluArgLysAspAsnAspLysSerProPheSerGlyAspGlyGlnProPheLys 168
Db 533 CTCTAGTGTAGGGGGATATGATGTCCTCTTCAAGTGAAGATGACACACTTTATG 592
OY 169 AspLysProGluLysGlyGluAlaThrGlyProAspLeuGlu-----GlyLysAsp 185
Db 593 CACACTCCCAACAGAGAGGTGCTGTGATCTGATCTGAAGCTCAGCTGCTCACCCT 652
OY 186 IleGlnThrGlyPheAlaGlyProSerGluAlaGluSerThrHisLysAspThrLysLys 205
Db 653 GTG-----TCAGGCTCCAGCAATGTCGAGTTCGAGTTCGACCACACAGCAATGGA 700
OY 206 ProGlyTyrAsnGluLysLysProGluArgLysLysLysLysLysLysLysLysLys 225
Db 701 CTGGGCTCTAATGATGATCCCAAGGAGAAAGTCCATAGCGGTGCTTGCACACAGA 760
OY 226 AspLysThrAlaLysGluAlaAspAlaValAspValSerLeuValGluLysSerAsnAsp 245
Db 761 GGAATACTCGCGAGGGGCGAGTTCGCGGATGTGAGCTGTGGAGCGAGCAATGGA 820
OY 246 IleMetGlySerThrAsnPheLysGluLeuProGlyArgGluGlnLysAsnArgValAspAla 265
Db 821 ATCAGCGGACAGCAAAATTTAGGAGCTCCCTGGAAGAAAGAAAGAAAGAGTCATGCC 880
OY 266 GlySerGlnAsnAlaHisGlnLysValAlaPheHisLysSerProAlaProSerLys 285
Db 881 AGCAGCCAAATGCTCATCAAGAAAGATTAAGATTCTACTCCCAACAGCCCTCAAAA 940
OY 286 GlnLysArgLysGluLysSerSerAspAlaAlaLysSerThrAsnTyrAsnGluLeuPro 305
Db 941 GGAAGGTAAAGGGGGGAGCGAGGAGCAACAGGAAAGCGGTTCACATGAATGCC 1000
OY 306 LysAsnGlyLysGlySerThrArgLysGlyValAspHisSerAsnArgAsnGlnAlaThr 325
Db 1001 AAGACGAGCAAGGGCGGCTAGCAAGATGCCGAAGATTAAGGAGCAACCAATACC 1060
OY 326 LeuAsnGluLysGlnArgPheProSerLysGlyLysSerGlnLysLeuProLysProSer 345
Db 1061 TTGACTGAAAGCCAAAGGTTCCAGCAAGCAAGCAAGCCACTCT-----TCT 1108
OY 346 ArgGlyLeuAspAsnGluLeuLysAsnGluMetAspSerPheAsnGlyProSerHisGlu 365
Db 1109 CACAGCTTGTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 1168
OY 366 AsnIle-----IleThrHisGlyArgLysTyrHisTyrValProHisArgLysAsnAsnSer 384
Db 1169 GCGATTCCATATGACACAGACAGCAAGCAAGCCAC-----CCT 1204
OY 385 ThrArgAsnLysGlyMetProGlnLysGlySerTyr-----GlyArgLysProHisSer 403
Db 1205 ACAGGGATAGGGGATGTACACAGCGAGAGCTCTCGGGCTCGAGAGAGCCCACTCCC 1264
OY 404 AsnArgArgPheSerSerArgArgAspAspSerSerSerLysSerSerSerSerSer 423
Db 1265 CACCGCGCGAAGCACCCCAAGAA-----GACAGTAGTAGTATCATCATCATGAGGAGT 1321
OY 424 SerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 430

```

```

Db 1322 TCTAGCGAGAGCTGTGAC 1342
RESULT 15
AF260922 1655 bp mRNA linear ROD 21-NOV-2000
LOCUS Rattus norvegicus osteoregulin mRNA, complete cds.
DEFINITION AF260922
ACCESSION AF260922
VERSION GI:11244771
KEYWORDS
SOURCE Rattus norvegicus.
ORGANISM Rattus norvegicus.
REFERENCE 1 (bases 1 to 1655)
AUTHORS Petersen,D.N., Tkalcic,G.T., Mansolf,A.L., Rivera-Gonzalez,R. and
Brown,T.A.
TITLE Identification of osteoblast/osteocyte factor 45 (OF45), a
bone-specific cDNA encoding an RGD-containing protein that is
highly expressed in osteoblasts and osteocytes
J. Biol. Chem. 275 (46), 36172-36180 (2000)
MEDLINE 20549633
PUBMED 10967096
REFERENCE 2 (bases 1 to 1655)
AUTHORS Brown,T.A.
TITLE Direct Submission
JOURNAL Submitted (26-APR-2000) Cardiovascular and Metabolic Diseases,
Pfizer Inc, Eastern Point Road, Groton, CT 06340, USA
FEATURES
source location/Qualifiers
1..1655
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
62..1369
/note="extracellular matrix protein; expression highly
restricted to bone tissue; highly expressed in osteoblasts
and osteocytes"
/codon_start=1
/product="osteoregulin"
/db_xref="gi:11244772"
/db_xref="id="AAG33366.1"
/translatability="NO"
/translation="MKNVSVGLFLESTYMAAPKINEDSGSGGNCNHLASVKEPMV
GKTEBGRDAPHLIDONROGATILRNITOPKSLVTEPEVSDPKRKROSVLSVL
PFDVHNTNDYSEDTEMOORLDLONSFGOSKHTPARSKSTHYLTLPQIRITLSPED
SASPDLYRGDNDVPEFSGDGHFMHTPDGGAVSDEPSAGHVSSESVETVDPH
TNGIGSNEIIGREGHIGAYATRGTAQAGSADYSLVGSNEITGSTRFKRLPEKEG
NRDASNSOMAHOGKVEFYPOAPSKKVGSGSRHTRGAGYNEIKPSKSGASKDAEE
SKNOVTLTESORFPEKGGQSHSGSHSIEDSSNSLSREGIAIAHRTSHPTRN
KMSORGRSMASHRPHRVRVSRKODSESSSSSSSSSSSSSD"
BASE COUNT 503 a 396 c 444 g 312 t
ORIGIN
Alignment Scores: 126e-45 length: 1655
Pred. No.: 211 matches: 211
Score: 938.50 conservative: 45
Percent Similarity: 59.95% mismatches: 146
Best Local Similarity: 49.41% indels: 25
Query Match: 41.18% gaps: 8
DB:
US-09-700-696b-2 (1-430) x AF260922 (1-1655)
OY 9 AsnLysGluAsnThrHisAsnGlyLeuArgMetSerIleTyrProLysSerThrGlyAsn 28
Db 146 AACCAAGCAACATCCAC-----TTAGCATCTGTGAAGCCTGAGCCCATGTGGGT 196
OY 29 LysGlyPheGluAspGlyAspAspAlaLysSerLysLeuHisAspGlnGluTyrGly 48
Db 197 AAGAACAGAGAGGTGGGAGATCTCCCTTACACCTGTTGACCAACAGCAGGAGT 256
OY 49 AlaAlaLeuIleArgAsnMetGlnHisIleMetGlyProValThrAlaIleLysLeu 68
Db 257 GCCACCCCTCTCGAATAATATCATCTACCTAGCCCTGTAAGAGTGTGAGCGGAGCTGAAGTA 316

```

```

OY 69 LeuGIuGIuAsnLysGIuAsnThrProArgAsnValLeuAsnIleIleProAlaSer 88
Db 317 CAGAGCCGACAGAAACAAAGAGAGAAAGACCTCAGAGTGTCTTACGTAATTCACACAGAT 376
OY 89 MetAsnTyrAlaLysAlaHisSerLysAspLysLysLysProGlnArgAspSerGlnAla 108
Db 377 GTCTCACATACTAAACGACTACTCAGAAAGATACAGAAACCAACAGAGAGATCTACTCTC 436
OY 109 GlnLysSerProValLysSerLysSerThrHisArgIleGlnHisAsnIleAspTyrLeu 128
Db 437 CAGAAACAGCCAGACAAACAGCAACACCCCTCGGGCCGACCAACAGCAGCACTACTACTA 496
OY 129 LysHisLeuSerLysValLysLysIleProSerAspPheGluGluSerGlyTyrThrAsp 148
Db 497 ACACATCTCCCCCAATCAGAAAGATCTCAGTACCTTCAGAGACAGACAGTCTCCCAAC 556
OY 149 LeuGlnGluArgAspAsnAspIleSerProPheSerGlyAspGluInProPheLys 168
Db 557 CTCTACTAGAGGGGATTAATGATGTCCTCTTCAGTGAGATGACAACTTTATG 616
OY 169 AspIleProGlyLysGluAlaThrGlyProAspLeuGlu-----GlyLysAsp 185
Db 617 CACACTCCCGACAGAGAGTGTGTGATGATCTGAAAGCTCAGCTGTGACCT 676
OY 186 IleGlnThrGlyPheAlaGlyProSerGluAlaGluSerThrHisLeuAspThrLysLys 205
Db 677 GTG-----TCAGGCTCCACCAATGTCAGATGTTGACCCACACAGAAATGA 724
OY 206 ProGlyTyrAsnGluIleProGluArgGluAsnGlyGlyAsnThrIleGlyThrArg 225
Db 725 CTGGGCTTAATGATGATCCAGGAGAGAGAGTCAATAGCGGCTTCATGCACACAGA 784
OY 226 AspGluThrAlaLysGluAlaAspAlaValAspValSerLeuValGluLysSerAsnAsp 245
Db 785 GGAATAAATCGCCAGCGGCGAGTTCGGGATGTGAGCTTGTGAGGCGACCAATGA 844
OY 246 IleMetGlySerThrAspPheLysGluLeuProGlyArgGluGlyAsnArgValAspAla 265
Db 845 ATCAGCGGCGAGTACCAATTTAGGAGCTCCCTGGAAAGAAAGAAACAGAGTCATGCC 904
OY 266 GlySerGlnAsnAlaHisGlnLysValGluPheHisTyrProProAlaProSerLys 285
Db 905 AGCACCACAAATGCTCATCAAGAAAGTAGAATTTCACTACCCACAGACCCCTCAAAA 964
OY 286 GluLysArgLysGluLysSerSerAspAlaAlaGluSerThrAsnTyrAsnGluIlePro 305
Db 965 GAGAAAGTAAAGGGGCGAGCAGAGCACACAGGAAAGCCGGTTACAAATGAATCCCC 1024
OY 306 LysAsnGlyLysGlySerThrArgLysGlyValAspHisSerAsnArgAsnGlnAlaThr 325
Db 1025 AAGAGCACCAAGGGCGGCTAGCAGAGATCCGGAACATCTAAAGGAAACCAAGTAACC 1084
OY 326 LeuAsnGluLysGlnArgPheProSerLysGlyLysSerGlnGlyLeuProIleProSer 345
Db 1085 TTGACTGAAGACCAAGAGTTCCAGGCCAAAGGCCAAAGCCAGTCT-----TCT 1132
OY 346 ArgGlyLeuAspAsnGluIleLysAsnGluMetAspSerPheAsnGlyProSerHisGlu 365
Db 1133 CACAGTCTTGATATGAGTTAAAGTAAGTAAGACTCTTCTAATAGTCTCAGTAGAGAG 1192
OY 366 AsnIle---IleThrHisGlyArgLysTyrHisTyrValProHisArgGlnAsnAsnSer 384
Db 1193 GCGATTTCATATACACACAGAGAACACAGCCAC-----CCT 1228
OY 385 ThrArgAsnLysGlyMetProGlnGlyLysGlySerTyr---GlyArgGlnProHisSer 403
Db 1229 ACACGGAAATAGGGGATGTCACAGCGAGAGGCTCTTGGGCTTCGAGAAAGACCCATCCC 1288
OY 404 AsnArgArgPheSerSerArgArgArgAspSerSerGluSerSerAspSerGlySer 423
Db 1289 CACCGGCGGTAGACACCGCCCAAGA---GACAGTAGTAGATCATCATCAGTGGAGCT 1345

```

```

OY 424 SerSerGluSerAspGlyAsp 430
Db 1346 TCTAGCGAGAGCAGTGTGAC 1366

```

Search completed: April 17, 2003, 01:36:55
Job time : 4024 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: April 16, 2003, 17:38:10 ; Search time 325 Seconds

(without alignments)
2979.567 Million cell updates/sec

Title: us-09-700-696b-2

2279

Sequence: 1 VNKEYSISNKNENTHNGLRMS.....RRDSSSESSDSSSESDGD 430

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 218539 segs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+ p2n model -DEV=xlh
-DB=N-Geneseq.101002 -OPMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -DOOPT=0
-LOOPT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blonsum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09700696.ecgn.1.1.263 &runat.11042003.101724.12376 -NCPU=6 -ICPU=3
-NO_XIPXY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV=TIMEOUT=120
-NARN.TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database : N-Geneseq.101002.*

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2279	100.0	1655	21	AAZ36447	cDNA encoding a po
2	2279	100.0	1655	22	AAE83759	Truncated phosphat
3	2276	99.9	1575	24	ABA99159	Human phosphatoin
4	2276	99.9	1662	24	ABA99160	Human phosphatoin
5	2276	99.9	1876	22	AAH26810	Human osteoregulin
6	2276	99.9	1869	22	AAH26811	Human osteoregulin
7	2276	99.9	2013	22	AAH83764	Human osteoregulin
8	2276	99.9	2019	22	AAH26808	Phosphatoin polyp
9	2276	99.9	2112	22	AAH26809	Human osteoregulin
10	2248	98.6	2031	22	AAH26807	Human osteoregulin
11	983.5	43.2	1682	22	AAH26807	Human osteoregulin
12	938.5	41.2	1655	22	AAH26807	Human osteoregulin
13	551.5	24.2	807	22	AAH26806	Human osteoregulin
14	551.5	24.2	807	22	AAH26806	Human osteoregulin
15	165	7.2	5688	23	AAK93391	Human cDNA clone r
16	165	7.2	5688	23	ABL26543	Drosophila melanog
17	150.5	6.6	2511	23	AAH26542	Drosophila melanog
18	150	6.6	6755	19	AAV21511	Human cDNA clone r
19	150	6.6	8577	23	ABV23751	Drosophila melanog
20	150	6.6	9295	23	ABL28667	Drosophila melanog
21	150	6.6	12294	23	ABL28666	Drosophila melanog
22	150	6.6	21314	23	ABL2494	Drosophila melanog
23	148.5	6.5	8201	21	AAH8864	Human dentin stalo
24	148.5	6.5	8201	21	ABO73537	Human dentin stalo
25	148	6.5	2652	23	AAH26543	Drosophila melanog
26	148	6.5	4409	23	ABV23751	Drosophila melanog
27	148	6.5	4409	23	ABV26878	Human prostate exp
28	148	6.5	4409	23	ABV26878	Human prostate exp
29	148	6.5	4409	23	ABV26878	Human prostate exp
30	146	6.4	1430	21	AAH26543	Drosophila melanog
31	144.5	6.3	1998	21	AAH26543	Drosophila melanog
32	142.5	6.3	1430	21	AAH26543	Drosophila melanog
33	142.5	6.3	1430	21	AAH26543	Drosophila melanog
34	141.5	6.2	1472	21	AAH26543	Drosophila melanog
35	141.5	6.2	1472	21	AAH26543	Drosophila melanog
36	141.5	6.2	1472	21	AAH26543	Drosophila melanog
37	141.5	6.2	1472	21	AAH26543	Drosophila melanog
38	140	6.1	3099	12	AAO13114	Human RNA polymera
39	140	6.1	3099	12	AAO13114	Human RNA polymera
40	139.5	6.1	1719	21	AAH26543	Drosophila melanog
41	139	6.1	35465	22	AAH26543	Drosophila melanog
42	138.5	6.1	2384	14	AAO46543	Drosophila melanog
43	138.5	6.1	2384	14	AAO46543	Drosophila melanog
44	138.5	6.1	2461	23	ABL08367	Drosophila melanog
45	138	6.1	3562	23	ABL08367	Drosophila melanog

ALIGNMENTS

RESULT 1

AAZ36447 standard; cDNA; 1655 BP.

AAZ36447;

22-FEB-2000 (first entry)

cDNA encoding a polypeptide designated phosphatoin.

Human; phosphatoin; Metastatic-tumour Excreted phosphatoin-Element;
MEPE; Na+-dependent phosphatoin cotransport; Vitamin D metabolism;
bone mineralisation; phosphate metabolism related disease;
hyperphosphatemia; renal osteodystrophy; renal dialysis;
secondary hyperparathyroidism; osteitis fibrosa cystica; hypercalcaemia;
X-linked hypophosphatemic rickets; hereditary hypophosphatemic rickets;
hypomeralised bone lesion; stunted growth; cystic fibrosis;
oncogenic hypophosphatemia osteomalacia; renal phosphate leakage;

KM renal osteodystrophy; osteoporosis; vitamin D resistant rickets;
 KW end organ resistance; renal Fanconi syndrome; autosomal rickets;
 KM Paget's disease; kidney failure; renal tubular acidosis; sprue; ds.
 XX Homo sapiens.
 OS
 XX
 XX Location/Qualifiers
 FT 1..1293
 FT /*tag= a
 FT /*product= "phosphatonin"
 FT /note= "5' end of the sequence is not given"
 FT 1615..1620
 FT /*tag= b
 FT polyA_signal
 XX
 XX W09960017-A2.
 XX
 XX 25-NOV-1999.
 XX
 XX 18-MAY-1999; 99WO-EP03403.
 XX
 XX 18-MAY-1998; 98GB-0010681.
 XX 04-SEP-1998; 98GB-0019387.
 XX
 XX (UNLO) UNIV COLLEGE LONDON.
 XX
 XX Rowe P:
 XX
 XX WPI: 2000-05362/04.
 XX DR P-PSDB; AAY53812.
 XX
 XX New polypeptides involved in the regulation of phosphate metabolism
 PT useful for diagnosing and treating disorders related to phosphate
 PT metabolism
 PT
 XX
 XX Claim 6; Fig 8; 136pp; English.
 XX
 XX The present sequence encodes a phosphatonin polypeptide (also called
 CC Metastatic-tumour Excreted Phosphaturic-Element (MEPE)). The level of
 CC phosphatonin in a subject modulates Na+-dependent phosphate cotransport,
 CC vitamin D metabolism and/or bone mineralisation. The phosphatonin
 CC polypeptides, polynucleotides, vectors and antibodies are used to treat
 CC phosphate metabolism related disease. They are used for treatment of
 CC hyperphosphatemia, or renal osteodystrophy, hyperphosphatemia in renal
 CC dialysis/pre-dialysis, secondary hyperparathyroidism or osteitis fibrosa
 CC cystica, or X-linked hypophosphatemic rickets, hereditary
 CC hypophosphatemic rickets with hypercalcaemia (HHRH), hypomineralised
 CC bone lesions, stunted growth in juveniles, oncogenic hyperphosphatemic
 CC osteomalacia, renal phosphate leakage, renal osteodystrophy,
 CC osteoporosis, vitamin D resistant rickets, end organ resistance, renal
 CC Fanconi syndrome, autosomal rickets, Paget's disease, kidney failure,
 CC renal tubular acidosis, cystic fibrosis or sprue. The polypeptide
 CC may also be used to manufacture combined preparations for simultaneous
 CC separate or sequential use for the treatment of phosphate metabolism
 CC disorders. A transformed osteoblast or bone cell line capable of
 CC phosphatonin overexpression is useful for the production of
 CC phosphatonin.
 CC
 XX
 XX Sequence 1655 BP; 609 A; 303 C; 380 G; 363 T; 0 other;
 SQ
 XX
 XX Alignment Scores:
 Pred. No.: 2.78e-177 Length: 1655
 Score: 2279.00 Matches: 430
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 21 Gaps: 0
 US-09-700-696b-2 (1-430) x AA26447 (1-1655)
 QY 1 Valasnlysglytyrserilserasnlysgluasnthrhisasnlyglueatrgmetser 20
 DB 1 GTGAATTAAGATATAGATACAGTAACAAAGAAATCTCAATGCGCTGAGATGTCA 60

QY 21 lleTyrProlyserThrGlyAsnlysglyphegluaspGlyAspAlaIleSerLys 40
 DB 61 AATTATCTTAATCTCAACGAGGATTAAGGTTTGAGATGGAGATGATGATACAGCAAA 120
 QY 41 LeuHISAspGlnGluGlyTyrGlyAlaIleuIleargAsnAsnMetGlnHISileMet 60
 DB 121 CTACATGACCAAGAAAGATATATGCGCAGCTCTCATACAGAAATAACATGCAACATTAATG 180
 QY 61 GlyProvalThrAlaIleuSerLeuGlyGlyGluAsnlysgluasnThrProArgAsn 80
 DB 181 GGCCCGTACGTCGATTAACCTCGGGGAGAGAAACAAAGAAACACACCTAGCAAT 240
 QY 81 ValIeuAsnIleIleProAlaSerMetAsnTyrAlaIleHisSerLysAspLysLys 100
 DB 241 GTTCTTAACATATCCACAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
 QY 101 LysProGlnArgAspSerGlnAlaGlnLysSerProValLysSerThrHisArg 120
 DB 301 AAGCTCAAAAGAGATCCCAAGCCAGAAAGTCCAGTAAGAAAGCAAGCAAGCCATGCT 360
 QY 121 IleGlnHisAsnIleAspTyrLeuLysHisLysSerLysValLysLysIleProSerAsp 140
 DB 361 ATTCAACACCAACATGACTACCTAAACATCTCTCAAAAGTCAAAATAATCCCAAGTCAAT 420
 QY 141 PheGluGlySerGlyTyrThrAspLeuGlnGluArgGlyAspAsnAspIleSerProPhe 160
 DB 421 TTTGAAGCGAGCGTTATACAGATCTTCAAGAGAGGAGACATGATATCTCTCTTC 480
 QY 161 SerGlyAspGlyGlnProPheLysAspIleProGlyLysGlyGluAlaThrGlyProAsp 180
 DB 481 AGTGGGAGCGCCCAACCTTTAAGACATTCCTGTAAGAGAAACCTACTGCTGAC 540
 QY 181 LeuGluGlyLysAspIleGlnThrGlyPheAlaGlyProSerGluAlaGlySerThrHis 200
 DB 541 CTACAGCAAGATATTCAACAGCGTTGACGCCCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 600
 QY 201 LeuAspThrLysLysProGlyTyrAsnGluIleProGluArgGluAsnGlyLysAsn 220
 DB 601 CTTGACACAAAGAGCCAGGTTATATGACATCTCCAGAGAGAGAGAAATGGTGGAAAT 660
 QY 221 ThrIleGlyThrArgAspGluThrAlaLysGluAlaAspAlaValAspValSerLeuVal 240
 DB 661 ACCATTGGAATAGGATGAAAGTCCGAAGAGCAAGTCTGTGATGTCACCTGTGA 720
 QY 241 GluGlySerAsnAspIleMetGlySerThrAsnProGlyGluLeuProGlyArgGluGly 260
 DB 721 GAGGCGACCAACGATATCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780
 QY 261 AsnArgValAspAlaGlySerGlnAsnAlaHisGlnGlyLysValGluPheHisTyrPro 280
 DB 781 AACACAGTGGATCTCGCAGCCAAAGTCCACCAAGGAGGATGATGATGATGATGATGATGAT 840
 QY 281 ProAlaProSerLysGluLysArgLysGluGlySerSerAspAlaIleGlySerThrAsn 300
 DB 841 CTGACACCTCAAAAGAGAAAGAAAGAGAGCAATGATGATGATGATGATGATGATGATGAT 900
 QY 301 TyrAsnGluIleProLysAsnGlyLysGlySerThrArgLysGlyValAspHisSerAsn 320
 DB 901 TATTAATGAATCTCTTAATAATGCGCAAGCGATACCAAGAGGATGATGATGATGATGAT 960
 QY 321 ArgAsnGlnAlaThrLeuAsnGluLysGlnArgPheProSerLysGlyLysSerGlnGly 340
 DB 961 AGGAACCAAGCAACCTTAATGAAAGCAAGGTTCTTACGAGGAGGAGGAGGAGGAGGAGG 1020
 QY 341 LeuProIleProSerArgLysLeuAspAsnGluIleLysAsnGluMetAspSerPheAsn 360
 DB 1021 CTGCCCATCTCTCTGCTGCTCTTGTATGTAATGAAATCAAAAGCAATGATGATGATGAT 1080
 QY 361 GlyProSerHisGluAsnIleIleThrHisGlyArgGlyTyrHisTyrValProHisArg 380
 DB 1081 GGCCCGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
 QY 381 GlnAsnAsnSerThrArgAsnLysGlyMetProGlnGlyLysGlySerTyrPheLysGln 400

```

Db 1141 CAAATATCTACACGGAATAGGTATGCGACACAGGAAGGCTCTCGGGTAGCA 1200
Qy 401 ProHisSerAsnArgAlaGlyPheSerSerArgArgAspAspSerSerGluSerSerAsp 420
Db 1201 CCCCATTCCACACAGAGATTAGTCCCTAGAGAGGATGACAGTACTGATCATCTGAC 1260
Qy 421 SerGlySerSerSerGluSerAspGlyAsp 430
Db 1261 AGTGCAGTTCACACTGACAGCAGATGTGAC 1290

RESULT 2
AAF83759
ID AAF83759 standard; cDNA; 1655 BP.
AC AAF83759;
XX
XX
XX 23-JUL-2001 (first entry)
XX
XX Truncated phosphatonin polypeptide (truncated MEPE) encoding cDNA.
XX
XX Metastatic-tumour excreted phosphaturic element; MEPE: phosphatonin;
XX KW phosphate; vitamin-D; skeletal formation; mineralization; truncated;
XX KM osteopathic; antihypertensive; cytoskeletal; human; ss.
XX
XX Homo sapiens.
XX
XX
XX Key 1..1293 Location/Qualifiers
XX CDS /*tag= a
XX FT /product= "truncated MEPE"
XX
XX WO200132878-A2.
XX
XX 10-MAY-2001.
XX
XX 31-OCT-2000; 2000WO-EP10747.
XX
XX 04-NOV-1999; 99US-0434185.
XX PR 08-NOV-1999; 99GB-0026424.
XX
XX (UNLO ) UNITV COLLEGE LONDON.
XX
XX Rowe P:
XX
XX MPI: 2001-343487/36.
XX DR P-PSDB: AAB62659.
XX
XX
XX New phosphatonin polypeptide a regulator of phosphate metabolism, for
XX PT diagnosing and treating disorders of phosphate, vitamin-D metabolism,
XX PT skeletal formation e.g. osteoporosis, Paget's disease, gout
XX
XX Example 4: Fig 8; 135pp; English.
XX
XX The invention relates to a novel human protein, metastatic-tumour
XX CC excreted phosphaturic element (MEPE) or phosphatonin (modulator of
XX CC phosphate and vitamin-D metabolism). The phosphatonin polypeptides,
XX CC polynucleotides and specific antibodies are useful for treating a
XX CC disorder of phosphate or vitamin D metabolism, skeletal formation and
XX CC mineralization. Phosphatonins are used to treat hyperphosphatemia, renal
XX CC osteodystrophy, secondary hyperparathyroidism, osteitis fibrosa cystica
XX CC or gout. It is used to prepare a medicament for treating X-linked
XX CC hypophosphatemic rickets, hereditary hypophosphatemic rickets with
XX CC hypercalciuria (HHRH), hypominerallized bone lesions, stunted growth in
XX CC juveniles, oncogenic hypophosphatemic osteomalacia, renal phosphate
XX CC leakage, renal osteodystrophy, osteoporosis, vitamin-D resistant rickets,
XX CC end organ resistance, renal Fanconi syndrome, autosomal rickets, Paget's
XX CC disease, kidney failure, renal tubular acidosis, cystic fibrosis or
XX CC sprue. Phosphatonin polynucleotides are useful as molecular weight
XX CC markers on Southern gels, as diagnostic probes for detecting the presence
XX CC of a specific mRNA. Phosphatonin polypeptides are also useful for
XX CC identifying agonists and antagonists, compounds which bind to
XX CC phosphatonin and drug candidates for therapy of phosphate metabolism

```

```

CC disorders. The present sequence represents the nucleotide sequence of a
CC first cloned cDNA for a truncated form of phosphatonin (MEPE).
XX
XX Sequence 1655 BP; 609 A; 303 C; 380 G; 363 T; 0 other:
XX
XX Alignment Scores:
XX Pred. No.: 2,78e-177 Length: 1655
XX Score: 2279.00 Matches: 430
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 100.00% Indels: 0
XX DB: 22 Gaps: 0
XX
XX US-09-700-696b-2 (1-430) x AAF83759 (1-1655)
XX
XX 1 ValAsnLysGluTyrSerIleSerAsnLysGluAsnThrHisAsnGlyLeuArgMetSer 20
XX Db 1 GTGAAATMAAGATATAGTATCTAGTAAACAGAAAGTACTCACAAATGGCTGAGATGTCA 60
XX
XX 21 IleTyrProLysSerThrGlyAsnLysGlyPheGluAspGlyAspAlaIleSerLys 40
XX Db 61 ATTTATCTTACTAGTCAACTGGGAATTAAGGCTTTGAGATGAGATGATGATCAGCAAA 120
XX
XX 41 LeuHisAspGlnGluGluTyrGlyAlaAlaLeuIleArgAsnAsnMetGlnHisIleMet 60
XX Db 121 CTACATGACCAAGAAATATGCGCAGCTCTCTCATCAAAATTAACATGCAATATATATG 180
XX
XX 61 GlyProValThrAlaIleLysLeuLeuGlyLysGluAsnLysGluAsnThrProArgAsn 80
XX Db 181 GGGCAGTGCAGCTCGCATTAACCTCTGGGGGAAAGAAACAAAGAACACACCTAGGAT 240
XX
XX 81 ValLeuAsnIleIleProAlaSerMetCAsnTyrAlaLysAlaHisSerLysAspLys 100
XX Db 241 GTTCTAAACATTAATCCACGAAATGATGATTAATCTTAAAGCACACTGGAAGATTAAG 300
XX
XX 101 LysProGlnArgAspSerGlnAlaGlnLysSerProValLysSerLysSerThrHisArg 120
XX Db 301 AAGCCTCAAAAGATTTCCCAAGCCCAAGAAAGTCCAGTAAACCAAAACACCCACATCGT 360
XX
XX 121 IleGlnHisAsnIleAspTyrLeuLysHisLeuSerLysValLysLysIleProSerAsp 140
XX Db 361 ATTCACACACAACTGACTACTTAACATCTCTCAAAAGTCAAAATAATCCACAGAT 420
XX
XX 141 PheGluGlySerGlyTyrThrAspLeuGlnGluArgGlyAspAsnAspIleSerProPhe 160
XX Db 421 TTTAAGGCAAGGCTTATACAGATCTTCAAGAGAGAGGAGCAATATATATCTCTTTC 480
XX
XX 161 SerGlyAspGlyLysProPheLysAspIleProGlyLysGlyLysAlaThrGlyProAsp 180
XX Db 481 AGTGGGACGCGCAACCTTTAAGGACATTCCTGGTAAAGAGACACTGCTCTGAC 540
XX
XX 181 LeuGluLysLysAspIleGlnThrGlyPheAlaGlyProSerGluAlaGluSerThrHis 200
XX Db 541 CTAAAGAGCAAAAGATATTTCAACAGAGGCTTTCAGGCCCAAGTAGAGCTGAGACTCAAT 600
XX
XX 201 LeuAspThrLysLysProGlyTyrAsnGlnIleProGluArgGluLysAsnGlyLysAsn 220
XX Db 601 CTTCACCAAAAAAGCCAGCTTTATATGATGCCACAGAGAGAAATAATGGTGGAAAT 660
XX
XX 221 ThrIleGlyThrArgAspGlnThrAlaLysGluAlaAspAlaValAspValSerLeuVal 240
XX Db 661 ACCATTGGAACTAGGATGAACCTGCGAAAGAGCAGAGATGCTGTATGACACCTTGA 720
XX
XX 241 GlnGlySerAsnAspIleMetGlySerThrAsnPheLysGluLeuProGlyArgGlyLys 260
XX Db 721 GAGGAGCAGCAGATATCAAGGTATGATACCAATTTTAAAGAGAGCTCCCTGAGAGAGA 780
XX
XX 261 AsnArgValAspAlaGlySerGlnAsnAlaHisGlnGlyValGluPheHisTyrPro 280
XX Db 781 AACAGAGTGGATCTGCGACCCAAATGCTCACAAAGGAGAGGTTGATTAATTAACCT 840
XX
XX 281 ProAlaProSerLysGluLysArgLysGlyLysSerSerAspAlaAlaGluSerThrAsn 300
XX

```


QY 221 ThrIleGlyThrArgAspGluThrAlaIysGluAlaIaspAlaIaspValSerLeuVal 240
 Db 946 ACCATTGGAACTAGAGGATGAAACTCCGAAGAGCGATGCTGTGATGATGACCTTGTA 1005
 QY 241 GluGlySerAspAspIleuMetGlySerThrAspPheGlyLeuProGlyArgGluGly 260
 Db 1006 GAGGCGACACAGATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1005
 QY 261 AsnArgValAspAlaGlySerGlnAsnAlaIleGlnGlyValAlaGluPheHisTyrPro 280
 Db 1066 AACGAGTGTGATGCTGCGACCAAAATGCTCAACAGGAAAGGTTGAGTTGATTCATTCCT 1125
 QY 281 ProAlaProSerGlyGlyValArgGlyGlySerSerAspAlaAlaGluSerThrAsn 300
 Db 1126 CTGCGACCTCTCAAAAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1185
 QY 301 TyrAsnGluIleProIysAsnGlyLysGlySerThrArgGlyValAlaPheHisSerAsn 320
 Db 1186 TATATGAAATTTCTAAATGCGAAAGCGAGTACAGAAAGGCTGATGATGATGATGATGAT 1245
 QY 321 ArgAsnGlnAlaIleThrLeuAsnGlyLysGlnArgPheProSerGlyLysSerGlnGly 340
 Db 1246 AGGAAACCAACACCTTAATGAAACAAAGGTTCTCTAGTAAAGGCAAAAGTCAAGGC 1305
 QY 341 LeuProIleProSerArgGlyLeuAspAsnGluIleLysAsnGluMetLysSerPheAsn 360
 Db 1306 CTGCGACCT 1365
 QY 361 GlyProSerHisGluAsnIleIleThrHisGlyArgLysTyrHisTyrAlaProHisArg 380
 Db 1366 GGGCCAGCATGATGAT 1425
 QY 381 GlnAsnAsnSerThrArgAsnGlyLysMetProGlnGlyLysGlySerTyrPheArgLys 400
 Db 1426 CAAATATATTTTACAGGATTAAGGTATGCCCAAGGAAAGGCTCCGCGGTAGACA 1485
 QY 401 ProHisSerAsnArgPheSerSerArgArgAspAspSerSerSerSerSerSerSer 420
 Db 1486 CCCCATTCACACAGAGGTTAGTCCCGTACAGAGATGACAGTATGATGATGATGATGAT 1545
 QY 421 SerGlySerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 430
 Db 1546 ACTGCGACGTTCAAGTGAAGAGCATGTGTGAC 1575
 RESULT 4
 ABA99160 standard; DNA; 1662 BP.
 XX ABA99160:
 AC 23-MAY-2002 (first entry)
 DT 23-MAY-2002 (first entry)
 DE Human phosphatoin encoding sequence including untranslated region.
 KW Human phosphatoin; cytosolic; antidiabetic; antiinflammatory;
 KW hyperphosphemia; arteriosclerosis; heart failure; gene;
 KW diabetic renal disease; kidney failure; cystic fibrosis; ds.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT CDS 8..1585 /tag= a
 FT /product= "Human phosphatoin"
 XX MO200198495-A1:
 XX 27-DEC-2001.
 XX 20-JUN-2001: 2001MO-JP05263.
 XX 21-JUN-2000: 2000CP-0191088.
 XX

PA (TAKE) TAKEDA CHEM IND LTD.
 XX Kurokawa T, Yamada T, Morimoto S;
 DR WPI: 2002-139791/18.
 DR P-PSDB; ABB08526.
 PT Phosphatoin of human origin and DNA encoding it for diagnosis and
 PT treatment of diseases associated with disorders of phosphate
 PT metabolism, e.g., hyperphosphemia, arteriosclerosis, heart failure,
 PT diabetic renal disease and kidney failure
 XX
 PS Claim 9: Page 123-124; 130pp; Japanese.
 CC This invention relates to human phosphatoin which has
 CC the activity of lowering blood phosphate and increasing urinary
 CC phosphate. The proteins are cytosolic, antidiabetic and
 CC antiinflammatory in their action. Phosphatoin down-regulates
 CC sodium-dependent phosphate transport in the kidney, it down-regulates
 CC 25-hydroxy-vitamin D3-1alpha-hydroxylase in the kidney and up-regulates
 CC useful in the diagnosis, treatment and prevention of phosphate
 CC metabolism related diseases such as hyperphosphemia, arteriosclerosis,
 CC heart failure, diabetic renal disease, kidney failure, acute coronary
 CC disease and cystic fibrosis. This sequence represents human
 CC phosphatoin encoding sequence and includes untranslated regions.
 XX
 S0 Sequence 1662 BP; 610 A; 317 C; 388 G; 347 T; 0 other;
 Alignment Scores:
 Pred. No.: 4,92e-177 Length: 1662
 Score: 2276.00 Matches: 429
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 99.77% Mismatches: 0
 Query Match: 99.87% Indels: 0
 DB: 24 Gaps: 0
 US-09-700-696b-2 (1-430) x ABA99160 (1-1662)
 QY 1 ValAsnIysGluThrSerIleSerAsnIysGluAsnThrHisAsnGlyLeuArgMetSer 20
 Db 293 CTGATATTAAGAAAT 352
 QY 21 IleTyrProIysSerThrArgAsnGlyLysPheGluAspGlyAspAlaIleSerLys 40
 Db 353 ATTTATCTTACGATCACTGGAATTAAGGTTGAGATGAGATGATGATGATGATGATGATGAT 412
 QY 41 LeuHisAspGlnGluIleThrGlyAlaAlaLeuIleArgAsnAsnMetGlnHisIleMet 60
 Db 413 CTACATGACCAAGAAAGATATGCGCGACGCTCTCATCAGAAATATATATATATATATATAT 472
 QY 61 GlyProValThrAlaIleLysLeuGluGlyGluAsnIysGluAsnThrProArgAsn 80
 Db 473 GGGCGAGTACGCGGATTAACCTCTGGGGGAAAGAAACAAAGAGACACACCTTAGAAT 532
 QY 81 ValLeuAsnIleIleProAlaSerMetAsnTyrAlaLysAlaHisSerLysAspLysLys 100
 Db 533 GTTCTTAACATATATCCACAGATATGATATATGCTTAAGCACACTCCAGATTAAGAAAG 592
 QY 101 LysProGlnArgAspSerGlnAlaGlnLysSerProValLysSerLysSerThrHisArg 120
 Db 593 AAGCTCAAGAGATTCACAGCCAGAAAGTCCAGTAAAGCAAAAGCAAGCAAGCAAGCAAG 652
 QY 121 IleGlnHisAsnIleAspTyrLeuLysHisLeuSerLysValLysLysIleProSerAsp 140
 Db 653 ATTCAACACAACTTGAATCACTTAACATCTCTCAAAAGTCAAAAGTCAAAAGTCAAAAGT 712
 QY 141 PheGluGlySerGlyTyrThrAspLeuGlnGluArgLysAspAsnAspIleSerProPhe 160
 Db 713 TTGAAGGACGCGGTTATACAGATCTTCAAGAGAGAGGAGCAATATATATATCTCTTTC 772
 QY 161 SerGlyAspGlyGlnProPheLysAspIleProGlyLysGlyGluAlaThrGlyProAsp 180
 |||||||

Db 478 GTTCTAAACATTAATCCAGCAAGATGATATGCTTAACACACCTCGAAGGATATAAAG 537
 Qy 101 LysProGlnArgAspSerGlnAlaGlnLysSerProValLysSerLysSerThrHisArg 120
 Db 538 AAGCTCAAAAGAGATCTCCCAAGCAAGAAAGCTCAAAAGCAAGCAAGCAAGCAAGCAAG 597
 Qy 121 IlleGlnHisasnIleAspTyrLeuLysHisLeuSerLysValLysLysIleProSerAsp 140
 Db 598 ATTCACACACAACTTGACTACCTAAACATCTCTCAAAAGCTCAAAAGCAAGCAAGCAAG 657
 Qy 141 PheGlnGlySerGlyTyrThrAspLeuGlnLysGlyAspAsnAspIleSerProPhe 160
 Db 658 TTGGAAGGACGGGTATACAGATCTTCAAGAGAGAGGACAAATGATATATCTCTTTC 717
 Qy 161 SerGlyAspGlyLysProPheLysAspIleProGlyLysGlyLysLysLysLysLys 180
 Db 718 AGTGGGAGCGGCAACCTTTTAAGGACATCTCTGTAAGAGAGAACTCTGCTCTGAC 777
 Qy 181 LeuGlnGlyLysAspIleGlnThrGlyPheAlaGlyProSerGlyValAlaGlySerThrHis 200
 Db 778 CTAGAAAGGCAAGATATCAACAGGCTTTCAGAGGCCCAAGTCAAGCTGAGTCTCAT 837
 Qy 201 LeuAspThrLysLysProGlyLysAsnGlnLysLysLysLysLysLysLysLysLys 220
 Db 838 CTGACACCAAAAGCCAGGTTATATGAGATCCAGAGAGAGAGAGAGAGAGAGAGAGAG 897
 Qy 221 ThrIleGlyThrArgAspLysLysLysLysLysLysLysLysLysLysLysLysLys 240
 Db 898 ACCATTGGAACTAGGAGTGAACCTCGAAAGAGGCAAGTCTCTGATGATGATGATGAT 957
 Qy 241 GlnGlySerAspAspIleGlnGlySerThrAspPheLysGlyLeuProGlyArgGlnGly 260
 Db 958 GAGGGCAGCAACGATATATGAGTATGACCAATTTTAAGAGAGAGAGAGAGAGAGAG 1017
 Qy 261 AsnArgValAspAlaGlySerGlnAsnAlaHisGlnGlyLysValGlnPheHisTyrPro 280
 Db 1018 AACAGAGTGAATGCTGGAGCAAAATGCTCACCAGGAGAGAGAGAGAGAGAGAGAG 1077
 Qy 281 ProAlaProSerLysGlnLysArgLysGlnLysSerSerAspAlaAlaGlySerThrAsn 300
 Db 1078 CTTGCACCTCAAAAG 1137
 Qy 301 TyrAsnGlnIleProLysAsnGlyLysGlySerThrArgLysGlyValAspHisSerAsn 320
 Db 1138 TATATGAAATTCCTAAATGCGCAAGGCAAGTCCGAAAGGAGTGTATGATCTTAAT 1197
 Qy 321 ArgAsnGlnAlaThrLeuAsnGlnLysGlnArgPheProSerLysGlyLysSerGlnGly 340
 Db 1198 AGGAACCAAGCAACCTTAATGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1257
 Qy 341 LeuProIleProSerArgGlyLeuAspAsnGlnLysLysAsnGlnLysLysSerPheAsn 360
 Db 1258 CTGCCCATCT 1317
 Qy 361 GlyProSerHisGlnAsnIleIleThrHisGlyArgLysTyrHisTyrValProHisArg 380
 Db 1318 GGGCCCGATATGAT 1377
 Qy 381 GlnAsnAsnSerThrArgAsnLysLysLysLysLysLysLysLysLysLysLysLys 400
 Db 1378 CAATAATTAATCTACAGCAATTAAGGATATGCAAGAGAGAGAGAGAGAGAGAGAG 1437
 Qy 401 ProHisSerAsnArgArgPheSerSerArgArgAspAspSerSerGlySerSerAsp 420
 Db 1438 CCCCATTCACAAAG 1497
 Qy 421 SerGlySerSerSerGlySerAspGlyAsp 430
 Db 1498 AGTGCAGTTCAAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1527
 RESULT 6
 AAH26811

ID AAH26811 standard; cDNA; 1969 bp.
 XX
 AC AAH26811;
 XX
 DT 21-DEC-2001 (first entry)
 XX
 DE Human osteoregulin (mature polypeptide) cDNA.
 XX
 KW Osteoregulin; human; bone; homeostasis; adipose; calcification;
 KW atherosclerosis; osteoporosis; osteopathic; antiarteriosclerotic;
 KW therapy; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT mat_peptide 1..1966
 FT /*tag= a
 XX
 PM EPI130098-A2.
 XX
 PD 05-SEP-2001.
 XX
 PF 27-FEB-2001; 2001EP-0301768.
 XX
 PR 29-FEB-2000; 2000US-185617P.
 PR 22-SEP-2000; 2000US-234500P.
 XX
 PA (PF12) PFIZER PROD INC.
 XX
 PI Brown TA, De Wet JR, Gowen LC, Hames LM;
 XX
 DR WPI; 2001-604111/69.
 XX
 PT P-PSDB; AAH82923.
 PT
 PS Novel osteoregulin polypeptide useful for regulating bone homeostasis,
 XX measuring the activity of osteoregulin.
 XX
 PS Claim 4: Page 58-59; 90pp; English.
 XX
 CC The present sequence is that of cDNA encoding human osteoregulin
 CC mature polypeptide (see AAH82923). It is derived from an
 CC osteoblast cDNA clone (see AAH26809), which encodes an osteoregulin
 CC polypeptide (see AAH82921) including an N-terminal signal peptide.
 CC Osteoregulin is a novel protein which plays a role in regulating bone
 CC homeostasis, adiposity, and the calcification of atherosclerotic
 CC plaques. A splice variant of this sequence (see AAH82922) was also
 CC identified. The invention provides osteoregulin proteins, nucleic
 CC acids encoding them, vectors, antibodies, host cells which express
 CC heterologous osteoregulin, and animal cells and mammals with a
 CC targeted disruption of an osteoregulin gene. The invention also
 CC provides screening assays to identify modulators of osteoregulin
 CC activity as well as methods of treating mammals for diseases or
 CC disorders associated with osteoregulin activity. The modulators of
 CC activity may be useful in the manufacture of a medicament for, as
 CC well as for treating, a mammal in need of regulation of bone mass
 CC and/or density, adiposity, atherosclerosis, vascular flexibility, and/or
 CC atherosclerotic plaque calcification (claimed), for treating and
 CC preventing osteoporosis, and for stimulating bone repair and
 CC regeneration.
 XX
 SQ Sequence 1969 BP; 732 A; 353 C; 447 G; 437 T; 0 other;
 Alignment Scores:
 Pred. NO.: 6.09e-177 Length: 1969
 Score: 2276.00 Matches: 429
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 99.77% Mismatches: 0
 Query Match: 99.87% Indels: 0
 DB: 22 Gaps: 0
 US-09-700-696b-2 (1-430) x AAH26811 (1-1969)

```

OY 1 ValAsnLysGluTyrSerIleSerAsnLysGluAsnThrHisAsnGlyLeuArgMetSer 20
DB 331 CCGAATAAAGAAATATATGATTCAGTACCAAGAGAAATACCAAAATGGCTGAGATGCA 390
OY 21 IleTyrProLysSerThrGlyAsnLysGlyPheGluAspGlyAspAspAlaIleSerLys 40
DB 391 ATTATACCTAGACCACTGGAAATGAAGGGTTTGAGAGAGGAGATGATCTATCCAGCAAA 450
OY 41 LeuHisAspGlnGluGluTyrGlyAlaAlaLeuIleArgAsnMetGlnHisIleMet 60
DB 451 CTACATGACCAAAAGAAATATGCGAGCTCTCATAGAAATATACATGCAACATTAATG 510
OY 61 GlyProValThrAlaIleLysLeuLeuGlyGluGluAsnLysGluAsnThrProArgAsn 80
DB 511 GGGCCAGTGCCTCGATTAACTCTGGGGAAGAAACAAAGAGAACACACTAGGAAT 570
OY 81 ValLeuAsnIleIleProAlaSerMetAsnTyrAlaLysAlaHisSerLysAspLysLys 100
DB 571 GTTCTAAACATATATCCAGCAAGTATGATTAATGCTTAAGCACACTCGAAGGATTAAGG 630
OY 101 LysProGlnArgAspSerGlnAlaGlnLysSerProValLysSerLysSerThrHisArg 120
DB 631 AAGCTTAAAGAGATTCCTCCAGCCAGCAAGAAAGTCCAGTAAAGCAAAAGCACCCATCGT 690
OY 121 IleGlnHisAsnIleAspTyrLeuLysHisLeuSerLysValLysLysIleProSerAsp 140
DB 691 ATTCACACCAACATGACTACTCTAAACATCTCTCAAAATCTCAAAATAATCCCACTGAT 750
OY 141 PheGluGlySerGlyTyrThrAspLeuGlnGluArgGlyAspAsnAspIleSerProPhe 160
DB 751 TTGTAAGGACAGCGTTATACAGATCTTCAAGAGAGAGGGGCAATGATATATCTCTTTC 810
OY 161 SerGlyAspGlyGlnProPheLysAspIleProGlyLysGlyGluAlaThrGlyProAsp 180
DB 811 AGTGGGACGGCCCAACTTTTAAGACATCTCTGTAAGAGAGAGTACTGCTGCTGAC 870
OY 181 LeuGluGlyLysAspIleGlnThrGlyPheAlaGlyProSerGluAlaGluSerThrHis 200
DB 871 CTAGAAAGCAAGATATTCACAAAGGGTTTTCAGAGCCCAATGAAACGTGAAGTACAT 930
OY 201 LeuAspThrLysLysProGlyTyrAsnGluIleProGluArgGluAsnGlyGlyAsn 220
DB 931 CTGACACAAAAAAGCCAGGTATATATGATCCAGAGAGAGAGAAAGAAATGGTGAAT 990
OY 221 ThrIleGlyThrArgAspGluThrAlaLysGluAlaAspAlaValAspValSerLeuVal 240
DB 991 ACCAATGGCACTAGGAGATGAACCTCGAAAGAGGCAATGCTGTTGATGATCAGCTTCTA 1050
OY 241 GluGlySerAsnAspIleMetGlySerThrAsnPhelyGluLeuProGlyArgGluGly 260
DB 1051 GAGGGACGACAGATATCTATGGGTAGTACCAATTTTAAGGAGCTCCCTCGAAGAGAACGA 1110
OY 261 AsnArgValAspAlaGlySerGlnAsnAlaHisGlnGlyLysValGluPheHisTyrPro 280
DB 1111 AACAGATGATGCTGGCAGCCAAATGCTCACAAAGGAGAAAGGTTGATTAATACCT 1170
OY 281 ProAlaProSerLysGluLysArgLysGluGlySerSerAspAlaAlaGluSerThrAsn 300
DB 1171 CCTGCACCTCAAAAGAGAAAGAAAGAGAGCAGTAGATGACAGCTGAAGTACCAAC 1230
OY 301 TyrAsnGluIleProLysAsnGlyLysGlySerThrArgLysGlyAlaLysPheSerAsn 320
DB 1231 TATAATGAATTCCTAAATAATGCAAGCAGTACAGAAAGGGTCTATGATCATCTTAAAT 1290
OY 321 ArgAsnGlnAlaThrLeuAsnGluLysGlnArgPheProSerLysGlyLysSerGlnGly 340
DB 1291 AGGAACCCAAACACTTAATGAAAAACAAGGTTTCTTCTAGTAAGGCAAAATGACGGGC 1350
OY 341 LeuProIleProSerArgGlyLysAspAsnGluIleLysAsnGluMetAspSerPheAsn 360
DB 1351 CTGCCCATCTCTCTCTGCTGCTTGAATGAATCAAAAGCAAAATGATCTCTTAAAT 1410
OY 361 GlyProSerHisGluAsnIleIleThrHisGlyArgLysTyrHisTyrValProHisArg 380

```

```

DB 1411 GGGCCAGTGCATGAGATATATATACACATGCCAGAAATATCATATATGACCCACAGA 1470
OY 381 GlnAsnAsnSerThrArgAsnLysGlyMetProGlnGlyLysGlySerTyrGlyArgGln 400
DB 1471 CAAATATATCTTACACGGAATTAAGGGTATGCCACAGAGGAAAGGCTCTGCTAGACAA 1530
OY 401 ProHisSerAsnArgArgPheSerSerArgArgAspAspSerSerGluSerSerAsp 420
DB 1531 CCCATTTCACACAGAGAGTTAGTTCCTGTAAGAAGGATGACAGTACTGATCATCTGAC 1590
OY 421 SerGlySerSerSerGluSerAspGlyAsp 430
DB 1591 AGTGCAGTTCACAGTGAAGGAGATGTCAC 1620

RESULT 7
AAE83764
ID AAE83764 standard; cDNA: 2013 BP.
AC AAE83764;
XX
XX 23-JUL-2001 (first entry)
DE Phosphatoinin polypeptide (MEPE) encoding cDNA.
XX
XX Metastatic-tumour excreted phosphaturic element; MEPE: phosphatonin;
KW phosphate; vitamin-D; skeletal formation; mineralization; truncated;
KW osteopathic; anti-gout; cytosolic; human; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 74..1651
XX FT /*tag= a
XX FT /product= "phosphatonin"
XX
XX PN MO200132878-A2.
XX PD 10-MAY-2001.
XX
XX 31-OCT-2000; 2000MO-BE10747.
XX
XX PR 04-NOV-1999; 99US-0434185.
XX PR 08-NOV-1999; 99GB-0026424.
XX
XX (UNLO ) UNIV COLLEGE LONDON.
XX
XX Rowe P;
XX WPI: 2001-343487/36.
XX DR P-PSDB; AAB62689.
XX
XX New phosphatonin polypeptide a regulator of phosphate metabolism, for
XX diagnosing and treating disorders of phosphate, vitamin-D metabolism,
XX skeletal formation e.g. osteoporosis, Paget's disease, gout
XX
XX Examples: Page 132-133; 135pp; English.
XX
XX The invention relates to a novel human protein, metastatic-tumour
XX excreted phosphaturic element (MEPE) or phosphatonin (modulator of
XX phosphate and vitamin-D metabolism). The phosphatonin polypeptides,
XX polynucleotides and specific antibodies are useful for treating a
XX disorder of phosphate or vitamin D metabolism, skeletal formation and
XX mineralization. Phosphatonins are used to treat hyperphosphatemia, renal
XX osteodystrophy, secondary hyperparathyroidism, osteitis fibrosa cystica
XX or gout. It is used to prepare a medicament for treating X-linked
XX hypophosphatemic rickets, hereditary hypophosphatemic rickets with
XX hypocalcemia (HHRH), hypomineralized bone lesions, stunted growth in
XX juveniles, oncogenic hypophosphatemic osteomalacia, renal phosphate
XX leakage, renal osteodystrophy, osteoporosis, vitamin D resistant rickets,
XX end organ resistance, renal Fanconi syndrome, autosomal rickets, Paget's
XX disease, kidney failure, renal tubular acidosis, cystic fibrosis or
XX sprue. Phosphatonin polynucleotides are useful as molecular weight

```

CC markers on Southern gels, as diagnostic probes for detecting the presence
 CC of a specific mRNA. Phosphatonic polypeptides are also useful for
 CC identifying agonists and antagonists, compounds which bind to
 CC phosphatonic and drug candidates for therapy of phosphate metabolism
 CC disorders. The present sequence represents the nucleotide sequence of a
 CC second cloned cDNA for the entire phosphatonic (MPPE).

XX Sequence 2013 BP: 740 A: 367 C: 459 G: 447 T: 0 other:

Alignment Scores:

Score: 6.22e-177 Length: 2013
 Percent Similarity: 2276.00 Matches: 429
 Best Local Similarity: 100.00% Conservative: 1
 Query Match: 99.77% Mismatches: 0
 Indels: 0
 Gaps: 0

US-09-700-696b-2 (1-430) x AAF83764 (1-2013)

OY 1 ValAsnLysGluTyrSerIleSerAsnLysGluAsnThrHisAsnGlyLeuArgMetSer 20
 Db 359 CTGATTAAGAAATAGTATCAGTAACAAAGAAATCTCAGATGCGCTGAGATCTCA 418
 OY 21 IleTyrProLysSerThrGlyAsnLysGlyPheGluAspGlyAspAlaIleSerLys 40
 Db 419 ATTATCTTAAGTCACTGGGAATTAAGGCTTTGAGAGATGAGATGATCTATCAGCAAA 478
 OY 41 LeuHisAspGlnGluGluTyrGlyAlaAlaLeuIleArgAsnAsnMetGlnHisIleMet 60
 Db 479 CTACATGACCAAGAAATATGGGCGAGCTCTCATCGAAATATCAATGCAACATTAATG 558
 OY 61 GlyProValThrAlaIleLysLeuLysGlyGluAsnLysGluAsnThrProArgAsn 80
 Db 539 GGGCGATGCTGAGTAAACCTCTGGGGGAAACCAACAAAGCAACCTAGGAAT 598
 OY 81 ValLeuAsnIleIleProAlaSerMetAsnTyrAlaLysAlaHisSerLysAspLysLys 100
 Db 599 GTTCTTAACATTAATCCAGCAACTATGAATATGCTAAACACACTCGAAGATTAAGAAG 658
 OY 101 LysProGlnArgAspSerGlnAlaGlnLysSerProValLysSerLysSerThrHisArg 120
 Db 659 AAGCCTCAAGAGATTCCTCCCAAGCCAGAAAGTCAGTAAAGCAACCCATCGT 718
 OY 121 IleGlnHisAsnIleAspTyrLeuLysHisLysSerLysValLysLysIleProSerAsp 140
 Db 719 ATTCAACACAACTGACTACCTAAACAATCTCTCAAAAGTCAAAATAATCCCTGAT 778
 OY 141 PheGlnGlySerGlyTyrThrAspLeuGlnArgGlyAspAsnAspIleSerProPhe 160
 Db 779 TTTGAAGGCAAGCGCTTATCAGATCTTCAAGAGAGAGGACATGATATCTCTTTC 838
 OY 161 SerGlyAspGlyGlnProPheLysAspIleProGlyLysGlyGluAlaThrGlyProAsp 180
 Db 839 AGTGGGAGCGCCACCTTTAAAGACATTCCTGGTAAAGAGAGACTGCTCTGAC 898
 OY 181 LeuGlnGlyLysAspIleGlnThrGlyPheAlaGlyProSerGlyAlaGlnSerThrHis 200
 Db 899 CTAGAAGGCCAAGATATCAACAGGGTTTGCAGGCCCAAGTGAAGCTGAGAGTACTCAT 958
 OY 201 LeuAspThrLysLysProGlyTyrAsnGluIleProGluArgGluGluAsnGlyLysAsn 220
 Db 959 CTTGACACAAAGAGCTTATATAGATCCAGAGAGAGAGAAATGTTGTAAT 1018
 OY 221 ThrIleGlyThrArgAspGluThrAlaLysGluAlaAspAlaValAspValSerLeuVal 240
 Db 1019 ACCATTGGAACTAGGAGAAACTCGAAAGAGCAGATGCTGTTGATGTCAGCCTTGT 1078
 OY 241 GluGlySerAsnAspIleMetGlySerThrAsnPheLysGluLeuProGlyArgGlnGly 260
 Db 1079 GAGGGCACAAGAGATATGAGTACCAATTTAAGAGAGCTCCCTGGAAGAAGAGA 1138
 OY 261 AsnArgValAspAlaGlySerGlnAsnAlaHisGlnGlyLysValGluPheHisTyrPro 280

Db 1139 AACAGATGATGCTGCGACCAAAATGCTCACCAAGGAAGGTGATGTTCTTACCT 1198
 OY 281 ProAlaProSerLysGluLysArgLysGlyGlySerSerAspAlaIleGlnSerThrAsn 300
 Db 1199 CTTGACCTCTCAAAAGAGAAAGAAAGAGAGGACGATGATGCTACCTGAAATACCAAC 1258
 OY 301 TyrAsnGluIleProLysAsnGlyLysGlySerThrArgLysGlyValAspHisSerAsn 320
 Db 1259 TATATGAATTCCTTAAATGCAAGGAGAGACAGACCAAGAGGCTTACATCTTCTAAT 1318
 OY 321 ArgAsnGlnAlaThrLeuAsnGlyLysGlnArgPheProSerLysGlyLysSerGlnGly 340
 Db 1319 AGSACCAACCAACCTTAATGAAAAACAAGGTTCTAGTAAGGCAAAAGTCAAGGC 1378
 OY 341 LeuProIleProSerArgLysLeuAspAsnGluIleLysAsnGlnMetLysPheAsn 360
 Db 1379 CTGCGCATTCCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1438
 OY 361 GlyProSerHisGluAsnIleIleThrHisGlyArgLysTyrHisTyrValProHisArg 380
 Db 1439 GGGCCGAGTCATGAGATTAATTAACACATGCGCAAAATATCATTAATGATACCCACAGA 1498
 OY 381 GlnAsnAsnSerThrArgAsnLysGlyMetProGlnGlyLysGlySerThrPheArgGln 400
 Db 1499 CAAATTAATTTCAACAGGATTAAGGTTGCGCAAGGAAAGGCTCTGGGTTAGACAA 1558
 OY 401 ProHisSerAsnArgArgPheSerSerArgArgAspAspSerSerGlySerSerAsp 420
 Db 1559 CCCCATTCACACAGAGAGTTAGTCCGTTAGAAAGGATGACAGTACATGATCTTAC 1618
 OY 421 SerGlySerSerSerGluSerAspLysAsp 430
 Db 1619 ACTGCGACTTCAAGTGAAGCGATGTCAC 1648

RESULT 8
 AAH26808
 ID AAH26808 standard; cDNA; 2019 BP.
 XX
 AC AAH26808;
 XX
 DT 21-DEC-2001 (first entry)
 XX
 DE Human osteoregulin cDNA.
 XX
 KW Osteoregulin; human; bone; homeostasis; adipose; calcification;
 KW atherosclerosis; osteoporosis; osteopathic; antiarteriosclerotic;
 KW therapy; ss.
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 96..1673
 FT /*tag= a
 FT sig_peptide 96..206
 FT /*tag= b
 FT mat_peptide 207..1670
 FT /*tag= c
 FN EPI130098-A2.
 XX
 PD 05-SEP-2001.
 XX
 XX 27-FEB-2001; 2001EP-0301768.
 PF 29-FEB-2000; 2000US-185617P.
 PR 22-SEP-2000; 2000US-234500P.
 XX
 PA (PFIZ) PFIZER PROD INC.
 XX
 PI Brown TA, De Wet JR, Gowen LC, Hames LM;
 XX
 DR MPI; 2001-604111/69.
 DR P-PDB; AAB82920.

XX Novel osteoregulin polypeptide useful for regulating bone homeostasis,
 PT adiposity and calcification of atherosclerotic plaques comprises
 PT measuring the activity of osteoregulin -

XX Claim 4; Page 44-45; 90pp; English.

XX The present sequence is that of cDNA encoding human osteoregulin (see
 CC AA82920), a novel protein which plays a role in regulating bone
 CC homeostasis, adiposity, and the calcification of atherosclerotic
 CC plaques. The cDNA was isolated from human osteoblast cDNA by
 CC PCR amplification using primers based on mouse osteoregulin cDNA
 CC and human genomic DNA sequences. A splice variant of this sequence
 CC was identified (see AA826809). The invention provides osteoregulin
 CC proteins, nucleic acids which encode them, vectors, antibodies,
 CC host cells which express heterologous osteoregulin, and animal
 CC cells and mammals with a targeted disruption of an osteoregulin
 CC gene. The invention also provides screening assays to identify
 CC modulators of osteoregulin activity as well as methods of treating
 CC mammals for diseases or disorders associated with osteoregulin
 CC activity. The modulators of activity may be useful in the
 CC manufacture of a medicament for, as well as for treating, a mammal
 CC in need of regulation of bone mass and/or density, adiposity,
 CC vascular flexibility, and/or atherosclerotic plaque calcification
 CC (claimed), for treating and preventing osteoporosis, and for
 CC stimulating bone repair and regeneration.

XX Sequence 2019 BP; 740 A; 371 C; 459 G; 449 T; 0 other;

Alignment Scores:

Pred. No.: 6,24e-177 Length: 2019
 Score: 2276.00 Matches: 429
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 99.77% Mismatches: 0
 Query Match: 99.87% Indels: 0
 DB: Gaps: 0

US-09-700-696B-2 (1-430) x AA826808 (1-2019)

QY 1 ValAsnLysGluTyrSerIleSerAsnLysGluAsnThrHisAsnGlyLeuArgMetSer 20
 DB 381 CTGAATTAAGAAATATAGTATCACTAACAAAGAAATATACACAAATGGCTGAGATGTA 440
 QY 21 IleTyrProLysSerThrGlyAsnLysGlyPheGluAspGlyAspAspAlaIleSerLys 40
 DB 441 ATTATTCCTTAAGCACTGGGAATTAAGGCTTTGAGAGATGAGATGCTATCAGCAAA 500
 QY 41 LeuHisAspGlnGluGluTyrGlyAlaAlaLeuIleArgAsnAsnMetGlnHisIleMet 60
 DB 501 CTACATGACCAAGAAAGATATGGCGACGCTCATCAGAAATGAATGCAATATATAATG 560
 QY 61 GlyProValThrAlaIleLysLeuLeuGlyGluGluAsnLysGluAsnThrProArgAsn 80
 DB 561 GGGCCAGTACTGCGATTAACTCTGGGGAGAAACCAAGAAAGAACACACCTAGGAAT 620
 QY 81 ValLeuAsnIleIleProAlaSerMetAsnTyrAlaLysAlaHisSerLysAspLysLys 100
 DB 621 GTTCTAAACATATCCACGAAGATATGCTTAACACACACCTGGAAGATAAAAG 680
 QY 101 LysProGlnArgAspSerGlnAlaGlnLysSerProValLysSerLysSerThrHisArg 120
 DB 681 AAGCCTCAAAAGATTCACCAAGCCCAAAAAGTCCACTAAAAAGCAAAAGCACCCATCGT 740
 QY 121 IleGlnHisAsnIleAspTyrLeuLysHisLeuSerLysValLysLysIleProSerAsp 140
 DB 741 ATTCAACCAACATTCGACTTAACATCTCTCAAAAAGTCAAAAATCCCAAGAT 800
 QY 141 PheGluGlySerGlyTyrThrAspLeuGlnGluArgGlyAspAsnAspIleSerProPhe 160
 DB 801 TTTCAGGACGACGGTTATACAGATCTTCAAGAGAGAGGAGCAATGATATATCTCTTC 860
 QY 161 SerGlyAspGlyGlnProPheLysAspIleProGlyLysGlyGluAlaIleThrGlyProAsp 180

DB 861 ACTGGGAGCGGCCAACCTTTTAAGACATTCCTGGTAAAGAGAACTACTGCTCTGAC 920
 QY 181 LeuGluGlyLysAspIleGlnThrGlyPheAlaGlyProSerGluAlaGlySerThrHis 200
 DB 921 CTGAGAGCAAGAAATATTCAAACAGGGTTGGCAGGCCCAAGTGAAGCTACTCAT 980
 QY 201 LeuAspThrLysLysProGlyTyrAsnGluIleProGluArgGluGluAsnGlyLysAsn 220
 DB 981 CTGACACCAAAAAGCCAGGTTTATATGAGATCCCGACAGAGAAAGAAATGCTGAAAT 1040
 QY 221 ThrIleGlyThrArgAspGluThrAlaLysGluAlaAspAlaValAspValSerLeuVal 240
 DB 1041 ACCATTGGAACTGAGGATGTAACCTCGGAAAAGGCGCAGATCGTTGATGTCAGCTTGTA 1100
 QY 241 GluGlySerAsnAspIleMetGlySerThrAsnPheLysGluLeuProGlyArgGluGly 260
 DB 1101 GAGGGCGACAGCATATCAATGAGGTAGACCAATTTTAAGAGAGCTCCCTGGAAAGAGA 1160
 QY 261 AsnArgValAspAlaGlySerGlnAsnAlaHisGlnGlyLysValGluPheHisTyrPro 280
 DB 1161 AACGAGATGATCTGGCAGCCAAATGCTCACCAAGGAGAGGTGATTCATTACCT 1220
 QY 281 ProAlaProSerLysGlyLysArgLysGlyLysSerSerAspAlaGlySerThrAsn 300
 DB 1221 CCTGACCCCTCAAAAAGAGAAAGAAAGAGGAGAGGAGTATGATGCTGAAAGTACCAAC 1280
 QY 301 TyrAsnGluIleProLysAsnGlyLysGlySerThrArgLysGlyValAspHisSerAsn 320
 DB 1281 TATATGAAATCTTAATAATGCAAAAGGCAAGTACCAAGGAGGTGATGATCTTAAT 1340
 QY 321 ArgAsnGlnAlaThrLeuAsnGluLysGlnArgPheProSerLysGlyLysSerGlnGly 340
 DB 1341 AGCAACCAAGCAACCTTAATGAAGAAACAAAGGTTCTCTGTAAGGCAAAAGTCAAGGC 1400
 QY 341 LeuProIleProSerArgGlyLysAspAsnGluIleLysAsnGluMetLysSerPheAsn 360
 DB 1401 CTGCCATTCCTCTCGTGGTGTGATATGAATCAAAAAGCAATGATCTCTTAAT 1460
 QY 361 GlyProSerHisGluAsnIleIleThrHisGlyArgLysTyrHisTyrValProHisArg 380
 DB 1461 GGCCCAATCATGAGAAATATATACATGCGCAAAAATATCATTTATGACCCACAGAGA 1520
 QY 381 GlnAsnAsnSerThrArgAsnLysGlyMetProGlnGlyLysGlySerThrPglArgGln 400
 DB 1521 CAAATTAATTTCTACACGGGAATTAAGGTATGCCACAAAGGAAAGGCTCTGCGGTAGCA 1580
 QY 401 ProHisSerAsnArgArgPheSerSerArgArgArgAspAspSerSerGluSerSerAsp 420
 DB 1581 CCCCATTCCAACAGAGAGGTTAGTCCCGTAGAAGGATGACAGTATGAGTCACTCTGAC 1640
 QY 421 SerGlySerSerSerGluSerAspLysAsp 430
 DB 1641 AGTGGCAGTTCAAGTAGAGAGCGATGTGAC 1670

RESULT 9
 AAH26809
 ID AAH26809 standard; cDNA: 2112 BP.
 AC AAH26809;
 XX 21-DEC-2001 (first entry)
 DT Human osteoregulin cDNA.
 XX
 DE
 XX Osteoregulin; human; bone; homeostasis; adipose; calcification;
 KW atherosclerosis; osteoporosis; osteopathnic; antiatherosclerotic;
 KW therapy; ss.
 XX Homo sapiens.
 OS
 XX key Location/Qualifiers
 FH key 96..1766
 FT CDS /tag= a

```

FT      sig_peptide      /transl_except= (pos:264..266,aa:Pro)
FT      96..206
FT      /*tag= b
FT      mat_peptide      207..1763
FT      /*tag= c
XX      EPI130098-A2.
XX
XX      05-SEP-2001.
XX
XX      27-FEB-2001; 2001EP-0301768.
XX
XX      29-FEB-2000; 2000US-185617P.
XX      22-SEP-2000; 2000US-234500P.
XX
XX      (PEIZ ) PEIZER PROD INC.
XX
XX      Brown TA, De Wet JR, Gowen LC, Hames LM,
XX      WPI: 2001-604111/69.
XX      P-PSDB: AAB82921.
XX
XX      Novel osteoregulin polypeptide useful for regulating bone homeostasis,
XX      PT      adiposity and calcification of atherosclerotic plaques comprises
XX      PT      measuring the activity of osteoregulin
XX
XX      Claim 4; Page 47; 90pp; English.
XX
XX      The present sequence is that of cDNA encoding human osteoregulin (see
XX      CC      AAB82920), a novel protein which plays a role in regulating bone
XX      CC      homeostasis, adiposity, and the calcification of atherosclerotic
XX      CC      plaques. The cDNA was isolated from human osteoblast cDNA by
XX      CC      PCR amplification using primers based on mouse osteoregulin cDNA
XX      CC      and human genomic DNA sequences. A splice variant of this sequence
XX      CC      was identified (see AAB82808). The invention provides osteoregulin
XX      CC      proteins, nucleic acids which encode them, vectors, antibodies
XX      CC      cells and mammals which express heterologous osteoregulin, and animal
XX      CC      gene. The invention also provides disruption of an osteoregulin
XX      CC      modulators of osteoregulin activity as well as methods of treating
XX      CC      mammals for diseases or disorders associated with osteoregulin
XX      CC      activity. The modulators of activity may be useful in the
XX      CC      manufacture of a medicament for, as well as for treating, a mammal
XX      CC      in need of regulation of bone mass and/or density, adiposity,
XX      CC      vascular flexibility, and/or atherosclerotic plaque calcification
XX      CC      (claimed), for treating and preventing osteoporosis, and for
XX      CC      stimulating bone repair and regeneration.
XX
XX      SQ      Sequence 2112 BP; 776 A; 383 C; 478 G; 475 T; 0 other;
XX
XX      Alignment Scores:
XX      Pred. No.:      6,66-177      Length:      2112
XX      Score:      2276.00      Matches:      429
XX      Percent Similarity:      100.00%
XX      Best Local Similarity:      99.77%
XX      Query Match:      99.87%      Mismatches:      0
XX      DB:      22      Indels:      0      Gaps:      0
XX
XX      US-09-700-696b-2 (1-430) x AAB26809 (1-2112)
XX
XX      1 ValAsnLysGluTyrSerIleSerAsnLysGluAsnThrHisAsnGlyLeuArgMetSer 20
XX      Db      CTGAATTAAGAAATATCTATCATGTAACCAAGAGAAATCTACACATGGCCGAGAGATGCA 533
XX
XX      21 IleTyrProLysSerThrGlyAsnLysGlyPheGluAspGlyAspAspAlaIleSerLys 40
XX      Db      ATTTATCTCTAAGTCAACTGGGAATTAAGGGTTTGAGAGATGGAATATGATCTATCACCATA 593
XX
XX      41 LeuHisAspGlnGluGlyTyrGlyAlaIleLeuIleArgAsnAsnMetGlnHisIleMet 60
XX      Db      CTGCATGACCAACAAAGAAATATGGCGACGCTCTCATGAAATTAACATGCAACATATTAATG 653
XX
XX      61 GlyProValThrAlaIleLysLeuGluGlyGluAsnLysGluAsnThrProArgAsn 80

```

```

Db      654 GGGCCAGTGAAGTGGATTAAATCTCTGGGGAGAGAAAAACAAGAGAACACCTAGGAAT 713
XX
XX      81 ValLeuAsnIleIleProAlaSerMetAspTyrAlaLysAlaHisSerLysAspLys 100
XX      Db      GTTCTAAACATTAATCCAGCAAGTATGAAATTAATGCTTAAGACACATCTGAAAGGATTAAG 773
XX
XX      101 LysProGlnArgAspSerGlnAlaGlnLysSerProValLysSerLysSerThrHisArg 120
XX      Db      774 AAGCCCTAAAGAGATTCACCAAGCCAGAAAAGTCCAGTAAAGCAAAAGCAACCCATCGT 833
XX
XX      121 IleGlnHisAsnIleAspTyrLeuLysHisLysSerLysValLysLysIleProSerAsp 140
XX      Db      ATTCACACACACATTTACTACTTAACATCTCTCAAAAGTCAAAAAGTCAAAAATCCCCAGTAT 893
XX
XX      141 PheGluGlySerGlyTyrThrAspLeuGlnGluArgGlyAspAspAspIleSerProPhe 160
XX      Db      894 TTTGAAGCGCAGCGTTATACAGATCTTCAAGAGACAGCGGCAATGATATCTCTCTTC 953
XX
XX      161 SerGlyAspGlyGlnProPheLysAspIleProGlyLysGlyGluAlaThrGlyProAsp 180
XX      Db      954 AGTGGGAGCGCCAACTTTTAAAGACATCTCTGTTAAAGGAGTAAGTGTGCTGAC 1013
XX
XX      181 LeuGluGlyLysAspIleGlnThrGlyPheAlaGlyProSerGluAlaGluSerThrHis 200
XX      Db      1014 CTGAAGGCAAGATATCAAAACAGGTTTGCAGGCCCAAGTGAAGCTGAGAGTACTCAT 1073
XX
XX      201 LeuAspThrLysLysProGlyTyrAsnGluIleProGluArgGluGluAsnGlyLysAsn 220
XX      Db      1074 CTGGACACAAAAGAACCGAGTTATTAATGATGCCAGACAGACAGAAAGTGTGGAAT 1133
XX
XX      221 ThrIleGlyThrArgAspGluThrAlaLysGluAlaAspAlaValAspValSerLeuVal 240
XX      Db      1134 ACCATTGGAACTAGGATGAAATCGCAAGAGGACATGCTGTGATGTCAGCTTGTGA 1193
XX
XX      241 GluLysSerAsnAspIleMetGlySerThrAspPheLysGluLeuProGlyArgGluGly 260
XX      Db      1194 GAGGCGACAAACGATATCATGAGGATGAGTACCAAAATTTTAAGCAGCTCCCTGGAAAGAAAGA 1253
XX
XX      261 AsnArgValAspAlaGlySerGlnAsnAlaHisGlnGlyLysValGluPheHisTyrPro 280
XX      Db      1254 AACGAGCTGATGCTGCGACCAAAATGCTCACCAAGGAGAGTTGAGTTCTTACCT 1313
XX
XX      281 ProAlaProSerLysGlyLysArgLysGluGlySerSerAspAlaAlaGluSerThrAsn 300
XX      Db      1314 CTGCAACCTCTCAAAAGAGAAAGAAAGAAAGAGCAGTATGATGATCAGCTCAAGTACCAC 1373
XX
XX      301 TyrAsnGluIleProLysAsnGlyLysGlySerThrArgLysGlyValAspHisSerAsn 320
XX      Db      1374 TATTAATGAATTCCTAAATAATGCAAGCAGCAGTACCAAGAGGATGATCATCTTAAT 1433
XX
XX      321 ArgAsnGlnAlaThrLeuAsnGlnLysGlnArgPheProSerLysGlyLysSerGlnGly 340
XX      Db      1434 AGGACCAAGCAACCTTAATATGAAAAACAAGGTTTCTTAAGGCAAAAGTCAAGGCG 1493
XX
XX      341 LeuProIleProSerArgGlyLeuAspAsnGluIleLysAsnGluMetLysSerPheAsn 360
XX      Db      1494 CTGCCCATTCCTCTCGTGTGTTGATTAATGAAATCAAAAACGAATGGAATCTCTTAAT 1553
XX
XX      361 GlyProSerHisGluAsnIleIleThrHisGlyArgLysTyrHisTyrValProHisArg 380
XX      Db      1554 GGGCCAGTCAAGAAATTAATTAACATGACAGCAAGAAATATCATTAATGATGCCACACGA 1613
XX
XX      381 GlnAsnAsnSerThrArgAsnLysGlyMetProGlnGlyLysGlySerThrGlyArgGln 400
XX      Db      1614 CAATAATATTCTACACGAAATTAAGGATATGCCACAAAGGAAAGCTCTCTGGGTAGACAA 1673
XX
XX      401 ProHisSerAsnArgArgPheSerSerArgArgArgAspAspSerSerGluSerSerAsp 420
XX      Db      1674 CCCCATTCACACAGAGGATTAATCCCGTGAAGGAGTACACATAGTGAATCTCTGAC 1733
XX
XX      421 SerGlySerSerSerGluSerAspGlyAsp 430

```

Db 1734 AGTGCAGTTCAAGTACAGAGCGATGTCATC 1763
 RESULT 10
 AAK94437
 ID AAK94437 standard; cDNA; 2031 BP.
 XX AAK94437;
 AC
 XX 06-NOV-2001 (first entry)
 DT
 XX Human full-length cDNA, SEQ ID NO: 3222.
 DE
 XX Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
 KW
 XX Homo sapiens.
 OS
 XX
 PN EP1130094-A2.
 XX
 PD 05-SEP-2001.
 XX
 PF 07-JUL-2000; 2000EP-0114089.
 XX
 PR 08-JUL-1999; 99JP-0194486.
 PR 11-JAN-2000; 2000JP-0118774.
 PR 02-MAY-2000; 2000JP-0183765.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 DR WPI: 2001-524255/58.
 DR P-PSDB: AAM93507.
 XX
 PT 830 Primers useful for synthesizing full length cDNA clones and their
 PT use in genetic manipulation -
 XX
 PS Claim 8; SEQ ID NO 3222; 1380bp + sequence listing; English.
 XX
 CC The invention relates to primers for synthesizing full length cDNA
 CC clones; 830 cDNA molecules encoding a human protein have been
 CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
 CC molecules have been determined. Primers for synthesizing the full length
 CC cDNA are useful for clarifying the function of the protein encoded by
 CC the cDNA. The full length clones were obtained by construction of full
 CC length enriched cDNA libraries that were synthesized by the oligo-capping
 CC method. The primers enable the production of the full length cDNA easily
 CC without any special methods. The present sequence is a full length
 CC human cDNA of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in CD-ROM format directly from EPO.
 CC
 XX
 XX Sequence 2031 BP; 740 A; 374 C; 463 G; 454 T; 0 other;
 SQ
 Alignment Scores:
 Pred. No.: 1,23e-174 Length: 2031
 Score: 2248.00 Matches: 427
 Percent Similarity: 99.77% Conservative: 2
 Best Local Similarity: 99.30% Mismatches: 1
 Query Match: 98.64% Indels: 1
 DB: 22 Gaps: 0
 US-09-700-696b-2 (1-430) x AAK94437 (1-2031)
 OY 1 ValAsnLysGluTyrSerIleSerAsnLysGluAsnThrHisAsnGlyLeuArgMetSer 20
 Db 372 CTGATTAAGAAATATAGTATCAGTACAAAGAAATCTCACAATGCGCTGAGAGATCTCA 431
 OY 21 IleTyrProLysSerThrGlyAsnLysGlyPheGluAspGlyAspAspAlaIleSerLys 40
 Db 432 ATTATATCTAGTCAACTGGGAATAAAGGCTTTGAGAGATGAGATGCTATAGCAAA 491
 OY 41 LeuHisAspGlnGluTyrGlyAlaIleLeuIleArgAsnAsnMetGlnHisIleMet 60

Db 492 CTACATACCAAGAAATATAGCGCACTCTCATCAGAAATATACATATCAATATATG 551
 OY 61 GlyProValThrAlaIleLysLeuGlyGluAsnLysGluAsnThrProArgAsn 80
 Db 552 GGGCCAGTACTGCGATTAACCTCTGGGGAGAAAAAACAAGAACACACCTAGGAAT 611
 OY 81 ValLeuAsnIleIleProAlaSerMetAsnTyrAlaLysAlaHisSerLysAspLysLys 100
 Db 612 GTTCTAAACATATCCACAGCAAGTATGATATCTTAAGCACACCTGAGAGATAAAAAA 671
 OY 101 LysProGlnArgAspSerGlnAlaGlnLysSerProValLysSerLysSerThrHisArg 120
 Db 672 AAGCTTCAAGAGATTCACAGACCAGCAAGAAAGTCCAGTAAAGCAAGACCCATCGT 731
 OY 121 IleGlnHisAsnIleAspTyrLeuLysHisLeuSerValLysLysIleProSerAsp 140
 Db 732 ATTCAACACACATGCTACTCTAAACATCTCTCAAAAGTCMAAAATAATCCCGCAT 791
 OY 141 PheGluLysSerGlyTyrThrAspLeuGlnGluArgGlyAspAsnAspIleSerProPhe 160
 Db 792 TTTGAAGGCGAGGTTATACAGATCTTCACAGAGAGGAGCAATGATATCTCTTC 851
 OY 161 SerGlyAspGlyGlnProPheLysAspIleProGlyLysGlyGluAlaThrGlyProAsp 180
 Db 852 AG-GGGAGCGGCCAACCTTTAAGACATCTCTGTTAAGAGAGAAAGCTACTGCTTGAC 910
 OY 181 LeuGluGlyLysAspIleGlnThrGlyPheAlaGlyProSerGluAlaGluSerThrHis 200
 Db 911 CTAGAAGGCAAGATATTCAACAGAGGTTGCAGGCCCAAGTGAAGTCAAGTCTCAT 970
 OY 201 LeuAspThrLysLysProGlyTyrAsnGluIleProGluArgGluAsnLysLysAsn 220
 Db 971 CTGACACAAAAAGCCAGATTATATGATGCCAGAACAGAAAGAAATAATGCTGAAT 1030
 OY 221 ThrIleGlyThrArgAspGluThrAlaLysGluAlaAspAlaValAspValSerLeuVal 240
 Db 1031 ACCATTGGAACTAGGATGAATGCGAAAGAGCGACATGCTATGATCTGACCTGTA 1090
 OY 241 GluGlySerAsnAspIleMetGlySerThrAspPheLysGluLeuProGlyArgGluGly 260
 Db 1091 GAGGCGACCAACGATATCAATGAGGTAGTACCAATTTTAAGACGCTCCCTGGAAAGAA 1150
 OY 261 AsnArgValAspAlaGlySerGlnAsnAlaHisGlnGlyLysValGluPheHisTyrPro 280
 Db 1151 AACAGATGATGCTGGCAGCAACAAATGCTCACCAAGGAGGTGAGTTCTATTCCT 1210
 OY 281 ProAlaProSerLysGluLysArgLysGlyLysSerSerAspAlaIleGluSerThrAsn 300
 Db 1211 CCTGCACCTTCMAAAGAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1270
 OY 301 TyrAsnGluIleProLysAsnGlyLysGlySerThrArgLysGlyValAspHisSerAsn 320
 Db 1271 TATATGTAATTCCTAAAAATGGCAAGGCCAGTACCAAGAGGCTGTATGATCTTAAT 1330
 OY 321 ArgAsnGlnAlaThrLeuAsnGluLysGlnArgPheProSerLysGlyLysSerGlnGly 340
 Db 1331 AGGAACCAAGCAACCTTTAATGAAACAAAGGTTCTCTATAGGCAAAAGACAGAGC 1390
 OY 341 LeuProIleProSerArgGlyLeuAspAsnGluIleLysAsnGluMetAspSerPheAsn 360
 Db 1391 CTGCCATTCCTCTTCGTCGTGTATGATTAATGAAATCAAAAGCAAGATGATCTTAAT 1450
 OY 361 GlyProSerHisGluAsnIleIleThrHisGlyArgGlyTyrHisTyrValProHisArg 380
 Db 1451 GGGCCAGTCAAGAGATTAATACACATGGCAAAATATCATTAATAGTACCCACAGAGA 1510
 OY 381 GlnAsnAsnSerThrArgAsnLysGlyMetProGlnGlyLysGlySerTyrGlyGln 400
 Db 1511 CAATAATTTCTACAGGATTAAGGATATGCCCAAGGAAAGGCTCTGGGTGAGAAA 1570
 OY 401 ProHisSerAsnArgArgPheSerSerArgArgArgAspAspSerSerGluSerSerAsp 420

Db 1571 CCCATTCCACAGAGGTTAGTCCCGTAGAAGGATGACAGTACATCTGAC 1630
 QY 421 SerGlySerSerSerGlySerAspGlyAsp 430
 Db 1631 AGTGCAGTTCAGTGAAGAGCGATGCTGAC 1660

RESULT 11
 AAH26807
 ID AAH26807 standard; cDNA; 1682 BP.
 AC AAH26807;
 XX
 DT 21-DEC-2001 (first entry)
 XX
 DE Mouse osteoregulin cDNA.
 XX
 KM Osteoregulin; mouse; bone; homeostasis; adipose; calcification;
 KM atherosclerosis; osteoporosis; osteopathic; antiarteriosclerotic;
 KM therapy; ss.
 XX
 OS Mus musculus.
 XX
 FH Key Location/Qualifiers
 FT CDS 67..1392
 FT /*tag= a
 XX
 PN EP1130098-f2.
 XX
 PD 05-SEP-2001.
 XX
 PF 27-FEB-2001; 2001EP-0301768.
 XX
 PR 29-FEB-2000; 2000US-185617P.
 PR 22-SEP-2000; 2000US-234500P.
 XX
 PA (PF12) PFIZER PROD INC.
 XX
 PI Brown TA, De Wet JR, Goven LC, Hames LM;
 XX
 DR WPI: 2001-604111/69.
 DR P-PSDB: AAB82919.
 XX
 PT Novel osteoregulin polypeptide useful for regulating bone homeostasis,
 PT adiposity and calcification of atherosclerotic plaques comprises
 PT measuring the activity of osteoregulin -
 XX
 PS Claim 4: Page 42-43: 90pp; English.
 XX
 CC The present sequence is that of cDNA encoding mouse osteoregulin (see
 CC AAB82919), a novel protein which plays a role in regulating bone
 CC homeostasis, adiposity, and the calcification of atherosclerotic
 CC plaques. The cDNA was isolated on the basis of hybridisation to
 CC mouse osteoregulin cDNA, as well as PCR and 3'RACE amplification to
 CC mouse tibia cDNA. The invention provides osteoregulin proteins,
 CC nucleic acids which encode them, vectors, antibodies, host cells
 CC which express heterologous osteoregulin, and animal cells and
 CC mammals with a targeted disruption of an osteoregulin gene. The
 CC invention also provides screening assays to identify modulators of
 CC osteoregulin activity as well as methods of treating mammals for
 CC diseases or disorders associated with osteoregulin activity. The
 CC modulators of activity may be useful in the manufacture of a
 CC medicament for, as well as for treating, a mammal in need of a
 CC regulation of bone mass and/or density, adiposity, vascular
 CC flexibility, and/or atherosclerotic plaque calcification (claimed),
 CC for treating and preventing osteoporosis, and for stimulating bone
 CC repair and regeneration.
 XX
 SO Sequence 1682 BP; 510 A; 401 C; 432 G; 339 T; 0 other;

Alignment Scores:
 Pred. No.: 3 1e-71
 Score: 983.50
 Percent Similarity: 62.56%

Length: 1662
 Matches: 217
 Conservative: 52

Best Local Similarity: 50.47% Mismatches: 134
 Query Match: 43.15% Indels: 27
 DB: 22 Gaps: 8

US-09-700-696B-2 (1-430) x AAH26807 (1-1682)

QY 6 SerIleSerAsnLysGluAsnThrHisAsnGlyLeuArgMetSerIleTyrProLysSer 25
 Db 166 AGTCGGCAATTCATCAAGACAGCATTCACAAAGACATTCGATCTGATCTATCC 225
 QY 26 ThrGlyAsnLysGlyPheGluAsnProLysAspAlaIleSerLysLeuHisAspGln 45
 Db 226 ACGGTGATGAAGCAGACAGATGCGAAGGTCTCTCTACCCGCTGGCCAGGAC 285
 QY 46 GluTyrGlyAlaIleLeuIleArgAsnMetGlnHisIleMetGlyProValThrAla 65
 Db 286 AGGTATGGTGGCTGCTCTCTCGAAATATCACGCGCTGTAAAGAGTGTAGTCTGG 345
 QY 66 IleLysLeuLeuGlyGluGluAsnLysGluAsnThrProArgAsnValLeuAsnIle 85
 Db 346 GCCGAACACTAGGAGGAGAAACCCAGAGACCTCAGAGTGTCTTAAGGTAT 405
 QY 86 ProIleSerMetAsnTyrAlaLysAlaHisSerLysAspLysLysProGlnArgAsp 105
 Db 406 CCAGCAGATGTCATGATGCTTAAGTCTCTTAAAGACATTAAGATCAAGAGATTAT 465
 QY 106 SerGlnAlaGlnLysSerProValLysSer-----LysSerThrHisArgIleGln 122
 Db 466 CTGCTAACCCAGACAGCCGCGTCAAAGAACACACACACACACCCGCGACCCGA 525
 QY 123 HisAsnIleAspTyrIleLysHisLeuSerLysValLysLysIleProSerAspPheGlu 142
 Db 526 CGAGCACTCACTACCTGACACATCTCCACAGATCAAGAAAGATCCACAGACTTGA 585
 QY 143 GlySerGlyTyrThrAspLeuGlnGluArgGlyAspAsnAspIleSerProPheSerGly 162
 Db 586 GGCAGTGGCTCCCAAGATCTTCTAGTGAAGGAGATATATATATGCTCCCTTACAGTGA 645
 QY 163 AspGlyGlnProPheLysAspIleProGlyLysGlyGluAla---ThrGlyProAspLeu 181
 Db 646 GATGGGCAACATTTATGACATCTCGCAAGAGAGAGCTGCTGGTCTGCTGAA--- 702
 QY 182 GluLysAspIleGlnThrGlyPheAlaGlyProSerGluAlaGluSerThrHisLeu 201
 Db 703 -----AGCTCACTACTCCGCCCTCTCAGGCTCCAGCAAACTGATTATTGACCA 756
 QY 202 AspThrLysLysProGlyTyrAsnGluIleProGluArgGluGlnGlyLysAsnThr 221
 Db 757 CATATGAGTGGACTAGCTCTATATGAGATCCCGGAGACAGACATGATGGAGTGGC 816
 QY 222 IleGlyThrArgAspGluThrAlaLysGluAlaAspAlaValAspValSerLeuValGlu 241
 Db 817 TATGCAACCAAGAACAAAGCTGACAGGGGCGCTGTCAGGTGGAGCCCTTGGGG 876
 QY 242 GlySerAsnAspIleMetGlySerThrAsnPheLysGluLeuProGlyArgGluGln 261
 Db 877 GGCAGCAATGAATATCACAGCAGACACCAATTCATGAGGAATCCCGGAAAGAAAGAAC 936
 QY 262 ArgValAspAlaGlySerGlnAsnAlaHisGlnGlyLysValGluPheHisTyrPro 281
 Db 937 AGAATTAAATCCCGCAGCAAAATGCTCATCAAGGAAAGTGAATTGATCATTCACACA 996
 QY 282 AlaProSerLysGluLysArgLysGluLysSerSerAspAlaIleGluSerThrAsnTyr 301
 Db 997 GTGGCTTCAGAGAAAGGTAAGGGGGGCTGGAGCATCAGGAGAGCT---GGTTAC 1053
 QY 302 AsnGluIleProLysAsnGlyLysGlySerThrArgLysGlyValAspHisSerAsnArg 321
 Db 1054 AACGAATCCCAAGAGCAGCAAGAGTACTCTACCAAAAGATCGAAAGATCCAAAGGG 1113
 QY 322 AsnGlnAlaThrLeuAsnGluLysGlnArgPheProSerLysGlyLysSerGlnGlyLeu 341
 Db 1114 AACCAATTAACTTACTGTCAGAGCAAGGATTTCCAGGTAAAGGCAAGGAGGAGCCCT 1173

OS Homo sapiens.
 XX EP1130094-A2.
 XX
 PD 05-SEP-2001.
 XX
 PF 07-JUL-2000; 2000EP-0114089.
 XX
 PR 08-JUL-1999; 99JP-0194486.
 PR 11-JAN-2000; 2000JP-0118774.
 PR 02-MAY-2000; 2000JP-0183765.
 XX
 PA (HELI-) HELIX RES. INST.
 XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 XX
 DR WPI; 2001-524255/58.
 XX
 PT 830 Primers useful for synthesizing full length cDNA clones and their
 PT use in genetic manipulation -
 XX
 PS Example 11; SEQ ID NO 1851; 1380bp + sequence listing; English.
 XX
 CC The invention relates to primers for synthesizing full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been
 CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
 CC molecules have been determined. Primers for synthesizing the full length
 CC cDNA are useful for clarifying the function of the protein encoded by
 CC the cDNA. The full length clones were obtained by construction of full
 CC length enriched cDNA libraries that were synthesised by the oligo-capping
 CC method. The primers enable the production of the full length cDNA easily
 CC without any special methods. The present sequence was used as the
 CC representative sequence from a human clone which was used in
 CC homology searches to identify the clone.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in CD-ROM format directly from EPO.
 CC
 XX
 SQ Sequence 807 BP; 315 A; 156 C; 163 G; 165 T; 8 other:
 XX
 Alignment Scores:
 Pred. No.: 2,896-36 Length: 807
 Score: 551.50 Matches: 124
 Percent Similarity: 87.67% Conservative: 4
 Best Local Similarity: 84.93% Mismatches: 10
 Query Match: 24.20% Indels: 8
 DB: Gaps: 1
 US-09-700-696B-2 (1-430) x AAK93391 (1-807)
 OY 1 valaasnlysglurtysertleaserasnlysgluasnthrhisasnlygluaurgmetsert 20
 DB 373 CTGAATAAAGAAATATAGTTCAGTAAACAAGAAATACACAAATGGCTGGAGATGCA 432
 OY 21 lletyprtylserthrglyasnlysglypheguaspolyaspaspalaileserlys 40
 DB 433 ATTATCCCTAAGCAACTGGGATAAAGGTTGAGAGATGAGATGATCTATCAGCAAA 492
 OY 41 leuHiaspoglnuglutyrglyalaaleuilearyasnsmetclnhsilemet 60
 DB 493 CTACATGACCAAGAAATATGGCCCACTCTCATCAGAAATACATGCAACATTAATG 552
 OY 61 -glypro-valthrralailtylsleuleuglygluasnlysgluasnthrproarg 80
 DB 553 GGGGCCAAGTACTGGATTAACCTCTGGGGGCAAGAAACAAAGAACACCTAGCN 612
 OY 80 snvalleuasnilleleproalasermetasnlyralaalsasertlyasp--ly 99
 DB 613 ATGTCTTAACATTAATCCGCAAGATATGCTAAAGCACACCTCMAAGGGGTAA 672
 OY 99 slslysploglnarqspserglnalagin-lysserproval--lysserlysert 118
 DB 673 AAGGAAGCTCTCAAGAGATCCCAAGCCAGAAAAAGTCCAANTAAAAAGCAAAAAAGCA 732

OY 118 hr-hisargilecgnhisasnilleasptyrleuylshisreu-serlysalyslysll 137
 DB 733 CCCATCGTNTTCAACACACANTGGCTANCTAAACATCTCCCAAAAGATCAAAAAAT 792
 OY 137 epro 138
 DB 793 CCCC 796
 RESULT 15
 ABL26543
 ID ABL26543 standard; DNA; 5688 BP.
 XX
 AC ABL26543;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 31102.
 XX
 KW Drosophila: developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ds.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI; 2001-656860/75.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Claim 1; SEQ ID NO 31102; 21pp + sequence listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16175) and the encoded proteins
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABBS7737-ABBS72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 SQ Sequence 5688 BP; 1925 A; 1719 C; 1147 G; 897 T; 0 other:
 XX
 Alignment Scores:
 Pred. No.: 0.00136 Length: 5688
 Score: 165.00 Matches: 74
 Percent Similarity: 37.19% Conservative: 74
 Best Local Similarity: 18.59% Mismatches: 168
 Query Match: 7.24% Indels: 82
 DB: Gaps: 16
 US-09-700-696B-2 (1-430) x ABL26543 (1-5688)
 OY 9 asnllysgluasnthrhisasnlygluaurgmetsertlelyrtylserthrglyasn 28
 DB 4669 AATGAGACGACCAACCAATTAACAAGCAAGAACACATCCAAAGCAAGATAGACACAC 4728

```
OY 29 LysGlyPheGluAspGlyAspAlaIleSerLysLeuHisAspGlnGluTyrGly 48
Db 4729 CAA-----ATTAACATCCAAAGCCAGAAATGAG----- 4755
OY 49 AlaAlaLeuIleArgAsnMetGlnHisIleMetGlyProValThrAlaIleLysIleu 68
Db 4756 -----AACAAACCAAAATACATCCAA----- 4779
OY 69 LeuGlyGluGluAsnLysGluAsnThrProArgAsnValLeuAsnIleIleProAlaSer 88
Db 4780 -----AGCAAGAAATCAGAAACAAACACAGAACACATC-----CAAAGT 4818
OY 89 MetAspTyrAlaLysAlaHis-----SerLysAspLys 100
Db 4819 ATGAATGAGAACAGCAACCAACAAACAGAACACATCCAAAGCAGAGTCAAAAC 4878
OY 101 LysProGlnArgAspSerGlnAlaGlnLysSerProValLysSerLysSerThrHisArg 120
Db 4879 AACATCGAAAGCAGAGAAATCAGAAACCAACCAACATCCAAAGCAGAGTCAAAAC 4938
OY 121 IleGlnHisAsnIleAspTyrLeuLysHisIleSerLysValLysLysIleProSerAsp 140
Db 4939 AACCGAGAACACATCCAAAGCAAGAAATCAGAAACAAACACAGAACACATCCAAAGCAAG 4998
OY 141 PheGlnGlySerGlyTyrThrAspLeuGlnGluArgGlyAspAsnAspIleSerProPhe 160
Db 4999 AAAGAAACCAACACAGAACACATCCAAAGCAAGAGTCAAAACAC----- 5046
OY 161 SerGlyAspGlyGlnProPheLysAspIleProGlyLysGlyGluAlaThrGlyProAsp 180
Db 5047 -----AACCAAGACATCCAAAGCAAGAAATGAGAAACACAGAACACAC 5088
OY 181 LeuGlnGlyLysAspIleGlnThrGlyPheAlaGlyProSerGlnAlaGlnSerThrHis 200
Db 5089 TTCCAAGCAGAAAT---CAGAACACATCCAAAGCAAGAAATCAGACAAACAGTACAC 5145
OY 201 LeuAspThrLysLysProGly-----TyrAsnGlnIleProGluArgGluGluAsnGly 218
Db 5146 ATCCAAAGCAAGATCAGAACACACCCAGACAGCATCCAAAGCAAAATCAGAACAC 5205
OY 219 Gly---AsnThrIleGlyThrArgAspGluThrAlaLysGlnAlaAspAlaValAspVal 237
Db 5206 AACCAAGAACATCCAAAGCAAGAAATCAGACATCCAAAGCAAGAAAT----- 5253
OY 238 SerLeuValGluLysSerAsnAspIle---MetGlySerThrAsnPheLysGluLeuPro 256
Db 5254 -----CAAAACAAACACAGATTAACATCCAAAGCAGAAAT-----CAG 5292
OY 257 GlyArgGlnGlyAsnArgValAspAlaGlySerGlnAsnAlaHisGlnGlyLysValGlu 276
Db 5293 AACCAACACAGAAATACATCCAAAGCAAGAAATCAGACACACAGAAATACATCCAA 5352
OY 277 PheHisTyrProProAlaProSerLysGluLysArgLysGluGlySerSerAspAlaIle 296
Db 5353 -----AGCAAGAAATCAGAAACAAACCAACAGAAATACATCCAAAGCAAGAAAT 5397
OY 297 GluSerThrAsnTyrAsnGlnIleProLysAsnGlyLysGlySerThrArgLysGlyVal 316
Db 5398 CAGAACACACAGAAATACATCCAAAGCAAGAAATCAGACACACAGAAATACATCCAA 5454
OY 317 AspHisSerAsnArgAsnGlnAlaThrLeuAsnGlnLysGlnArgPheProSerLysGly 336
Db 5455 CAGAACACACAGAAATACATCCAAAGCAAGAAATCAGACACACAGAAATACATCCAA 5514
OY 337 LysSerGlnGlyLeuProIleProSerArgLysLeuAspAsnGlnIleLysAsnGlnMet 356
Db 5515 CAGAACACACAGAAATACATCCAAAGCAAGAAATCAGACACACAG-----AACACATG 5568
OY 357 AspSerPheAsnGlyProSerHisGlnAsnIleIleThrHisGlyArgLysTyrHisTyr 376
Db 5569 CAAAGCAGAAATCAGAACAAACCAAGAACATC----- 5601
OY 377 ValProHisArgGlnAsnAsnSerThrArgAsnLysGlyMetProGlnGlyLys 394
```

Db 5602 ---CAAAGCAAGAAATCAGAACAAACCAAGCAAGAAATCAGACAGAAAG 5652

Search completed: April 17, 2003, 00:30:15
Job time : 364 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 16, 2003, 19:35:10 ; Search time 1920 Seconds

(without alignments)
3627.114 Million cell updates/sec

Title: US-09-700-696B-2
Perfect score: 2279
Sequence: 1 VNKEYSISKENTHNCRLMS.....RRDSESSSDSSSSSDSD 430

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame_plus_p2n.model -DEV=xlh
-O=/cgn2.1/USPTO/spool/US09700696/runat_11042003_101725_12413/app.query.fasta.1.583
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPPC=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-USER=US09700696 -GCGN.1.11906 -etunal_11042003_101725_12413 -NCPU=6 -ICPU=3
-NO.XLPXY -NO.MMAP -LARGOQUERY -NEG.SCORER=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
EST:
1: em_estha:*
2: em_esthum:*
3: em_estin:*
4: em_estm:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	372.5	16.3	571	17	A2647898
C 2	328.5	14.4	516	14	B0783292
C 3	234.5	10.3	594	14	B0780656
C 4	193	8.5	443	10	BB849948
C 5	160.5	7.0	502	10	BE107579
C 6	159.5	7.0	499	14	BQ782396
C 7	158.5	7.0	1708	12	BG262039
C 8	156	6.8	1584	12	BF381948
C 9	154.5	6.8	1376	12	BG165801
C 10	154	6.8	1337	12	BE907390
C 11	150.5	6.6	934	17	A2186441
C 12	150.5	6.6	1366	12	BG287570
C 13	150	6.6	1367	12	BE779053
C 14	150	6.6	2126	12	BG030344
C 15	149.5	6.6	1068	12	BG538300
C 16	149	6.5	1856	10	BE612548
C 17	148	6.5	1866	12	BF217713
C 18	147.5	6.5	1677	12	BF797713
C 19	147	6.5	1286	12	BF381914
C 20	147	6.5	1632	12	BF785846
C 21	146.5	6.4	1729	12	BF216092
C 22	146.5	6.4	3167	11	BC026849
C 23	146	6.4	1885	12	BG121849
C 24	146	6.4	1904	12	BF317483
C 25	145.5	6.4	1518	12	BG289468
C 26	145	6.4	1109	12	BG425865
C 27	145	6.4	1302	12	BF215082
C 28	145	6.4	1704	12	BF235932
C 29	145	6.4	1884	12	BF235932
C 30	144	6.3	814	17	AG071025
C 31	144	6.3	1510	13	B1519016
C 32	143.5	6.3	1004	12	BG847844
C 33	143.5	6.3	2920	13	BG913272
C 34	143	6.3	1441	12	BG741972
C 35	143	6.3	2343	12	BF345940
C 36	142.5	6.3	1456	12	BF037615
C 37	142.5	6.3	1573	12	BG403502
C 38	142	6.2	1460	12	AG080078
C 39	142	6.2	2061	12	BE783704
C 40	142	6.2	1258	12	BG294539
C 41	141.5	6.2	1681	12	BF214150
C 42	141.5	6.2	1874	12	BF159055
C 43	141.5	6.2	1874	12	BF341908
C 44	141	6.2	1326	12	BF689145
C 45	141	6.2	1326	12	BF689145

ALIGNMENTS

RESULT 1
A2647898/c
LOCUS
DEFINITION IM0514108R Mouse 10kb plasmid UGCM1 library Mus musculus genomic
VERSION A2647898
ACCESSION A2647898.1 GI:11779824
KEYWORDS
SOURCE GSS.
ORGANISM house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,
Irlam, H., Longacre, S., Mahmood, M., Meenen, E., Pedersen, T., Rellly

TITLE 'M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. Mouse and white genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
Institute of Utah Genome Center

FEATURES
source

BASE COUNT	111 a	144 c	152 g	103 l	1
ORIGIN					1

Alignment Scores:	
Pred. No.:	1,15e-26
Score:	372.20
Percent Similarity:	59.89%
Best Local Similarity:	46.15%
Query Match:	16.34%
DB:	17
Length:	57
Matches:	84
Conservative:	25
Identities:	66
Misses:	7
Gaps:	3

US-09-700-696B-2 (1-430) x AZ647898 (1-571)

```

Oy      20 SerIleYrProLySerThrIyAnuSclYpHeIuSpGlyAspAlaIleSer 39
      |||||:||||| ||| :||| ||||| |||
Db      566 TCTGTATCTCTATCCACAGCGTGATGAAAGCAGACGATGGGCAAGCTCTCCCTT 50
      |||||
Oy      40 LysLeuHisApGInGluIuYrGlyAlaAlaLeuIleArGAsnAsnMetGInHisIle 59
      |||||: ||||| ||||| ||||| ||| |||
Db      506 CAAACCGCGCTGGCAGACAGATGATGGTGGCTCCCTCCCAAAATATCACGACGCTGTA 44
      |||||
Oy      60 MetGlyProValThrAlaIleLysLeuLeuGInGluGluAsnLysGluAsnThrProArg 79
      ||||| ||||| :||| ||| ||||| |||
Db      446 AAGAGCTAGTACGACTGGGCGCGAACTACGAGAGGAAGAAACCGAGGAAAGACTCAG 38
      ||||| ||||| :||| ||| ||||| |||
Oy      80 AsnValLeuAsnIleIleProAlaSerMetAsnYrAlaLysAlaHisSerLysAspLys 95
      :|||:||||| ||||| ||||| |||||

```

Db	386	AGTGTCTAAGCGTAAATTCACAGACGATGTCATGATCTTAAAGTCTCTTAAAGAACAATA	327
QY	100	LysLysProGlnArgAspSerGlnIleAlaGlnLysSerProValLysSer-----Lys	116
Db	326	AAGATTCAGACGAGTATCTGTCTAACCCAGACGCCGGGTCAAAAACCAACACACCACAA	267
QY	117	SerThrHisArgIleGlnHisAsnIleAspTyrIleuLysHisIleuSerLysValLysLys	136
Db	266	CACACCCCCCAGACCCGACGAGCAGCTACTACTGACACATCTCCACACATCAACAAAG	207
QY	137	IleProSerAspPheGlnGlySerGlyThrHisAspLeuGlnIuArgGlyAspAspAsp	156
Db	206	ACTCCCACTGACCTTGAAGGACGACGTGGCTCCCCAGATCTCTTAGTAGAGGGGAGATAATGAT	147
QY	157	IleSerProPheSerGlyAspGlyGlnProPheLysAspIleProGlyLysGlyGluA	176
Db	146	GTCGCCCTTTCAGTAGGAGATGGGCAACATTTTATGACACATTTCTCTGSCAAAGGAGTGCT	87
QY	177	---ThrGlyProAspLeuGlnGlyLysAspIleGlnThrGlyPheAlaGlyProSerGlu	195
Db	86	GGGTGTGGTCTGAA-----AGTCACACTACTGCGCCCTCTCAGGCTCCAGCAAA	36
QY	196	AlaGlu	197
Db	35	GCTGAA	30

RESULT 2	716 bp	mRNA	linear	EST 26-JUL-2002
LOCUS	B0783292/c			
DEFINITION	B0783292			
	UI-R-F00-cpf-b-19-0-UI s1	UI-R-F00	Rattus	
	UI-R-F00-cpf-b-19-0-UI 3',	mRNA sequence.	norvegicus	CDNA clone
ACCESSION	B0783292			
VERSION	B0783292.1	GI:21991764		
KEYWORDS	EST.			
SOURCE	Norway rat.			
ORGANISM	Rattus norvegicus			
	Chordata; Craniata; Vertebrata; Euteleostomi;			

FEATURES
SOURCE

```

/dev_stage="Abul"
/abp_host="DH10B (Life Technologies) (T1 phage resistant)"
/abp_note="Vector: pT73-Pec (Pharmacia) with a modified
polylinker. Site_1: EcoR I; Site_2: Not I; UI-R-PF0 is a
subtracted cDNA library containing the following tissue(s)
: Normal cartilage and SR-JWS Tumor Line . The

```



```

||||| |||||
Db      322 GAGAGCAGTGTGCAC 308

RESULT 4
BB849948      443 bp      mRNA      linear      EST 26-NOV-2001
LOCUS      BB849948      RIKEN full-length enriched, adult inner ear Mus musculus
DEFINITION      cDNA clone F930104E24 5', mRNA sequence.
ACCESSION      BB849948
VERSION      BB849948.1      GI:17091402
KEYWORDS      EST.
SOURCE      house mouse.
ORGANISM      Mus musculus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS      Akimura,T., Arikawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
              Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T., Imotani,K., Ishii,
              Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
              Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.,
              Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
              Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,
              A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.,
              Watanuki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
              RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
              2001)
              Unpublished (2001)
JOURNAL      Contact: Yoshinide Hayashizaki
COMMENT      Laboratory for Genome Exploration Research Group, RIKEN Genomic
              Sciences Center(GSC), Yokohama Institute
              The Institute of Physical and Chemical Research (RIKEN)
              1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
              Tel: 81-45-503-9222
              Fax: 81-45-503-9216
              Email: genome-res@sc.riken.go.jp,
              URL: http://genome-gsc.riken.go.jp/
              Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,
              M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
              Normalization and subtraction of cap-trapper-selected cDNAs to
              prepare full-length cDNA libraries for rapid discovery of new
              genes. Genome Res. 10 (10), 1617-1630 (2000)
              Wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
              Watanuki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsunura,
              S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
              Hayashizaki,Y.
              RIKEN integrated sequence analysis (RISA) system--384-format
              sequencing pipeline with 384 multicapillary sequencer. Genome Res.
              10 (11), 1757-1771 (2000)
              Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,
              Y. and Hayashizaki,Y.
              Computer-based methods for the mouse full-length cDNA
              encyclopedia: real-time sequence clustering for construction of a
              nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
              Please visit our web site (http://genome-gsc.riken.go.jp) for
              further details.
              e mouse tissues.

FEATURES
Source
1. 443
   /organism="Mus musculus"
   /strain="C57BL/6J"
   /db_xref="taxon:10090"
   /clone="F930104E24"
   /clone_1lb="RIKEN full-length enriched, adult inner ear"
   /tissue_type="inner ear"
   /dev_stage="adult"
   /note="pooled tissues : (tissue_type=cerebellum,
   dev_stage=16 days neonate,
   (tissue_type=cerebellum, dev_stage=0 day neonate,
   sex=mixed), (tissue_type=hippocampus, dev_stage=adult,
   sex=male), (tissue_type=whole body, dev_stage=9 days
   embryo, sex=mixed), (tissue_type=lung, dev_stage=13 days
   embryo, sex=mixed)"

BASE COUNT      124 a      106 c      128 g      85 t

```

```

ORIGIN
Alignment Scores:      3.71e-09      Length:      443
Pred. No.:      193.00      Matches:      42
Score:      58.82%      Conservative:      18
Percent Similarity: *      41.18%      Mismatches:      42
Best Local Similarity:      8.47%      Indels:      0
Query Match:      10      Gaps:      0
DB:

US-09-700-696b-2 (1-430) x BB849948 (1-443)
OY      2      ASnlysglUtyrserlleserAsnlysgluAsnThrHisAnGlyLeuArgMetserlle 21
Db      118      AATGACAGACGACGCGCGGCAATTCACAGGACCTTGCCACCATCTGG 177
OY      22      TyrrpolsrThrGlyAsnlysglypneGluAspGlyAspAlalleSerlyLeu 41
Db      178      TATCCTGAGCCACGCGTGGTGAAGCAGAGGAGGCGCAGGTCTCTCTTCAACCG 237
OY      42      HisApGlnGluGlyrGlyAlaAlaLeuileArgAsnMetGlnHisileMetGly 61
Db      238      CCGGCCAGACAGCATGTGCTGCTGCCCTCCAGAAATATCAGAAAGCTGTAAGACT 297
OY      62      ProValThrAlalleLyLeuLeuGlyGluGluAsnlysgluAsnThrProArgAsnVal 81
Db      298      CTGCTGACTGGGCGCAACCTACGAGGAGGAAGCAAGCAAGCAAGCAAGCAAGCAAGT 357
OY      82      LeuAsnlelleProAlaserMetAsnTyralAlaHisserLyAspLyLyLy 101
Db      358      CTAAGCGTAATTCAGCAGATGTCAATGATGCTTAAGATCTCCCTAAAGACATTAAGAA 417
OY      102      ProGln 103
Db      418      CAAGAG 423

RESULT 5
BE107579/c      502 bp      mRNA      linear      EST 13-JUN-2000
LOCUS      BE107579      UI-R-CAO-aww-d-02-0-UI.s1 UI-R-CAO Rattus norvegicus cDNA clone
DEFINITION      UI-R-CAO-aww-d-02-0-UI 3', mRNA sequence.
ACCESSION      BE107579
VERSION      BE107579.1      GI:8499684
KEYWORDS      EST.
SOURCE      Norway rat.
ORGANISM      Rattus norvegicus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
              Rattus.
              1 (bases 1 to 502)
              Bonaldo,M.F., Lennon,G. and Soares,M.B.
              Normalization and subtraction: two approaches to facilitate gene
              discovery
              Genome Res. 6 (9), 791-806 (1996)
              97044477
              Contact: Soares, MB
              Program for Rat Gene Discovery and Mapping
              University of Iowa
              451 Eckstein Medical Research Building Iowa City, IA 52242, USA
              Tel: 319 335 8250
              Fax: 319 335 9565
              Email: mssoares@blue.weeg.uiowa.edu
              The sequence contained an oligo-dT track that was present in the
              oligonucleotide that was used to prime the synthesis of first
              strand cDNA and therefore this may represent a bonafide poly A
              tail. cDNA library Preparation: M.B. Soares Lab Clone distribution:
              clones will be available through Research Genetics (www.reagen.com)
              Seq primer: M13 Forward
              POLYA=Yes.

FEATURES
Source
1. 502
   /organism="Rattus norvegicus"
   /strain="Sprague-Dawley"

```

```

/db_xref="taxon:10116"
/clone="UI-R-CA0-aw-d-02-0-UI"
/clone.lib="UI-R-CA0"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; The UI-R-CA0
library is a subtracted library derived from the following
tissues: thalamus, cerebellum, hypothalamus, medulla, pons
hippocampus. For a detailed description of the library
from which this clone was derived, please visit our web
site at ratsec.eng.uiowa.edu. The subtraction has been
previously described in (Bonaldo, Lennon and Soares,
Genome Research 6:791-806, 1996)
TAG_SEQ=None found"
BASE COUNT      112 a      128 c      104 g      157 t      1 others
ORIGIN
Alignment Scores:
Pred. No.:      6.69e-06      Length:      502
Score:          160.50      Matches:      40
Percent Similarity: 63.01%      Conservative: 6
Best Local Similarity: 54.79%      Mismatches: 16
Query Match:    7.04%      Indels:      11
DB:              10      Gaps:          4

US-09-700-696b-2 (1-430) x BE107579 (1-502)
OY 360 AsnGlyProSerHisGluAsnIle---IleThrHisGlyArgLysTyrHisTyrValPro 378
Db 501 AATAGTCTCAGTAGACGGGATTGCATATGACACAGGAGACAGACCCAC-----451
OY 379 HisArgGlnAsnAsnSerThrArgAsnLysGlyMetProGlnGlyLysSerTrp---397
Db 450 -----CMTACACGGAATAGAGGGGATGTACACAGCGAGAGCCTCTGGGCC 406
OY 398 GLyArgGlnProHisSerAsnArgArgPheSerSerArgArgArgPheSerSerGlu 417
Db 405 TCGAGAGACCCCATCCCGCCGCGGTACAGCACCGCCCAAGA---GACAGTAGTGAG 349
OY 418 SerSerAspSerGlySerSerSerGluSerAspGlyAsp 430
Db 348 TCATCATCCAGTGGAGTTCTACCGAGACGACGTGTGAC 310

RESULT 6
LOCUS      B0782396      499 bp      mRNA      linear      EST 26-JUL-2002
DEFINITION UI-R-FPO-cpk-e-05-0-UI.s1 UI-R-FPO Rattus norvegicus CDNA clone
ACCESSION B0782396
VERSION B0782396.1 GI:21990868
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 499)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9365
Email: msoares@blue.weeg.uiowa.edu
Tissue Procurement: Jeff Stevens
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

```

```

FEATURES
Source
Location/Qualifiers
1..499
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone.lib="UI-R-FPO-cpk-e-05-0-UI"
/clone.lib="UI-R-FPO"
/tissue_type="Mixed tissues"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site_1: EcoR I; Site_2: Not I; UI-R-FPO is a
subtracted cDNA library containing the following tissue(s)
): Normal cartilage and SR-JWS Tumor Line. The
subtraction was made according to Bonaldo, Lennon and
Soares, Genome Research, 6:791-806, 1996. The
oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (drr)18 tail. The
sequence tags for these libraries are: CTAAATGACG,
CATCTCTGTA,
TAG_LIB=UI-R-FPO
TAG_RISUE=cartilage
TAG_SEQ=CTAAATGACG"
BASE COUNT      112 a      123 c      103 g      160 t      1 others
ORIGIN
Alignment Scores:
Pred. No.:      8.34e-06      Length:      499
Score:          159.50      Matches:      40
Percent Similarity: 63.01%      Conservative: 6
Best Local Similarity: 54.79%      Mismatches: 16
Query Match:    7.00%      Indels:      11
DB:              14      Gaps:          4

US-09-700-696b-2 (1-430) x B0782396 (1-499)
OY 360 AsnGlyProSerHisGluAsnIle---IleThrHisGlyArgLysTyrHisTyrValPro 378
Db 499 AATAGTCTCAGTAGACGGGATTGCATATGACACAGGAGACAGACCCAC-----449
OY 379 HisArgGlnAsnAsnSerThrArgAsnLysGlyMetProGlnGlyLysSerTrp---397
Db 448 -----CCTACACGGAATAGAGGGGATGTACACAGCGGAGAGAGCCTCTGGGCC 404
OY 398 GLyArgGlnProHisSerAsnArgArgPheSerSerArgArgArgPheSerSerGlu 417
Db 403 TCGAGAGACCCCATCCCGCCGCGGTACAGCACCGCCCAAGA---GACAGTAGTGAG 347
OY 418 SerSerAspSerGlySerSerSerGluSerAspGlyAsp 430
Db 346 TCATCATCCAGTGGAGTTCTACCGAGACGACGTGTGAC 308

RESULT 7
LOCUS      BG262039      1708 bp      mRNA      linear      EST 13-FEB-2001
DEFINITION 602373873F1 NIH_MGC_94 Mus musculus CDNA clone IMAGE:4481203 5',
ACCESSION BG262039
VERSION BG262039.1 GI:12771855
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1708)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)

```


ORGANISM Homo sapiens
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 1584)
 NIH-MGC http://mgi.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: c9apbs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: CLONTECH Laboratories, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
 Plate: LCM883 row: 1 column: 12
 High quality sequence step: 32.

FEATURES
 Source
 1. 1584
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4050203"
 /clone_1lb="NIH-MGC-56"
 /tissue_type="Primitive neuroectoderm"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: Brain; Vector: pDNR-LIB (Clontech); Site_1: SfiI (ggccgcctggcc); Site_2: SfiI (ggccgcctggcc); Double-stranded cDNA was prepared from cell line RNA. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCCATATGCGC-3' and 3' adaptor sequence: 5'-ATCTAGAGCGCGCGCGCGAGCATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

BASE COUNT 725 a 310 c 503 g 46 t

Alignment Scores:
 Pred. No.: 7.03e-05 Length: 1584
 Score: 156.00 Matches: 95
 Percent Similarity: 36.388 Conservative: 60
 Best Local Similarity: 22.308 Mismatches: 183
 Query Match: 6.85% Indels: 89
 DB: 12 Gaps: 14

US-09-700-696B-2 (1-430) x BF381948 (1-1584)

QY 3 LysGluTyrSerIleSerAsnLysGluAsnThrHisAsnGlyLeuArgMetSerIleTyr 22
 DB 405 GAGAGACAAACAAATCAAAAGAGGGGGAATGCTACCAATGACCAACAGCGGAGCGCTA 464

QY 23 ProLysSerThrGlyAsnLysGlyPheGluAspGlyAspAspAlaIleSerLysLeuHis 42
 DB 465 CCTCCCAACACGACGACGACCGCGGAGAA----- 497

QY 43 AspGlnGluGluTyrGlyAlaAlaLeuIleArgAsnAsnMetGlnHisIleMetGlyPro 62
 DB 498 -----ATMAAAGACAAATAATTGCGGAGCCCGCGGCGCGC 533

QY 63 ValThrAlaIleLysLeuLysGluGluGluAsnThrProArgAsnValLeu 82
 DB 534 TGGGGGGCAACAAA-----AAACACACACACACGACACACAAACAC 581

QY 83 AsnIleIleProAlaSerMetAsnTyrAlaLysAlaHisSerLysAspLysLysPro 102
 DB 582 AAAACACACACAAAC 635

QY 103 GlnArgAspSerGlnAlaGlnLysSerProValLysSerLysSerThrHisArgIleGln 122
 DB 636 CAAAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 695

QY 123 HisAsnIleAspTyrLeuLysHisLeuSerLysValLysIleProSerAspPheGlu 142
 DB 696 CACACACACACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 755

QY 143 GlySerGlyTyrThrAspLeuGlnArgGlyAspAsnAsp-----IleSerProPhe 160
 DB 756 AGGAGAGAAAAGACGAAAGAGAGAGTAAAGAGAAAGAAAGAAAGAGAGCGACAGC 815

QY 161 SerGlyAspGlyGlnProPheLysAspIleProGlyLysGlyGlnAlaThrLysProAsp 180
 DB 816 ACCGGTACGAGAAAAGAGCGCGGAGAGACAGAGAGAAAGACACAGAGAGAGAGAGAG 875

QY 181 LeuGluGlyLysAspIleGlnThrGlyPheAlaGlyProSerGlnAlaGluSerThrHis 200
 DB 876 AA-CACACACAAACCAAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 934

QY 201 LeuAspThrLysLysProGlyTyrAsnGluIleProGluAlaGluGlnLysAsn 220
 DB 935 GAGAGCAGCAGAGAAAG 988

QY 221 ThrIleGlyThrArgAspGlu---ThrAlaLysGlnAlaAspAlaValAspValSerLeu 239
 DB 989 -----GACACAAAGGAG 1033

QY 240 ValGluGlySerAsnAspIleMetCysSerThrAsnPheLysGluLeuProGlyArgGlu 259
 DB 1034 -----ACGCCACAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1054

QY 260 -----GlyAsnArgValAspAlaGlySerGlnAsnAlaHisGlnGlyLysValGlu 276
 DB 1055 GAGGAGACGGGAG 1102

QY 277 PheHisTyrProProAlaProSerLysGluLysArgLysGluLysSerSerAspAla-Al 296
 DB 1103 -----CGAGAGAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1138

QY 296 agLysSerThrAsnTyrAsnGluIleProLysAsnGlyLysGlySerThrArgLysGlyVal 316
 DB 1139 AGAGAGCGGAAACAGCGAG 1196

QY 316 IAspHisSerAsnArgAsnGlnAlaThrLeuAsnGluLysGlnArgPheProSerLysGlu 336
 DB 1197 -----GAAGCGGAG 1219

QY 336 YLysSerGlnLysLeuProIleProSerArgGlyLeuAspAsnGluIleLysAsnGluMet 356
 DB 1220 CGAAAGCGAGAGCGAGACTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1276

QY 356 TAspSerPheAsnGlyProSerHisGluAsnIleIleThrHisGlyLysGlyTyrHisTyr 376
 DB 1277 AGATGAGAGAAACAG 1328

QY 376 rValProHisArgGlnAsnAsnSerThrArgAsnLysGlyMetProGlnGlyLysGlyLys 396
 DB 1329 -----CAAG 1384

QY 396 rTyrGlyArgGlnProHisSerAsnArgArg-----PheSe 408
 DB 1385 GAACCAAG 1444

QY 408 rSerArgArgArgAsp 413
 DB 1445 GCGAAG 1460

RESULT 9
 BG165801 1376 bp mRNA linear EST 06-FEB-2001
 LOCUS 602344467F1 NIH_MGC_89 Homo sapiens cDNA clone IMAGE:4454417 5',
 DEFINITION mRNA sequence.
 ACCESSION BG165801
 VERSION BG165801.1 GI:12672504
 KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens
REFERENCE NIH-MGC <http://mgi.nci.nih.gov/>.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: AFCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLAM10246 row: f column: 18
High quality sequence start: 8
High quality sequence stop: 152.
Location/Qualifiers
1. 1376
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4454417"
/clone_lib="NIH-MGC-89"
/tissue_type="hypermelanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Kidney; Vector: pCMV-SPORT6; Site: 1; Notch; Site: 2; Salt; Cloned unidirectionally; Oligo-dT primed. Average insert size 1.3 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH-MGC library."

BASE COUNT 617 a 409 c 318 g 32 t

ORIGIN

Alignment Scores:
Pred. No.: 8.38e-05 Length: 1376
Score: 154.50 Matches: 95
Percent Similarity: 34.36% Conservative: 61
Best Local Similarity: 20.93% Mismatches: 163
Query Match: 6.78% Indels: 135
DB: Gaps: 20

US-09-700-696b-2 (1-430) x BG155801 (1-1376)

QY 4 GtUTyTserTleserAsnLysGluAsnThr-----HisAsn 15
Db 294 GACCAAGACAGCAAGAAACAAAGAAACCAACACACAGCAAGAAAGCAAGACAC 353
QY 16 GtyleuArGmetSerTleTyTProLysSerThrGlyAsn-LysGlyPheGluAspGlyAs 35
Db 354 GGCACAAACACCAACGTCGACCCACAGGAAACAGCAACACACGCGGCGTGA 413
QY 35 pAspAla-----LleSerLysLeuHisAspGlnGluGtUTyTgLyAlaAlaLe 51
Db 414 GAACACAAACGCGACATGCGCAACAAACACACCCGCAAC-----453
QY 51 ulleArGsnAsnMetGlnHisLleMetGlyProValThrAlaLleLysLeuGlyG1 71
Db 454 ----GCCACTACCAAGCGCAAGACGCAACAAACGACCAATCAAA-----GA 500
QY 71 uGluAsnLysGluAsnThrProArGAsnValLeuAsnLleLleProAlaSerMetAsnTy 91
Db 501 GCAGCAACAGCAACACACACCAACCAACAAAGACGACGCGACCCACCCCC 560
QY 91 rAlaLysAlaHisSerLysAspLysLysProGlnArGAspSerGlnAlaGlnLysSe 111
Db 561 AACCCGCGCAACACCAACCGCAAGCAACGAAGAAAGCCTAACAAACCAACAGAGC 620
QY 111 rProValLysSerLysSerThrHisArGtlleGlnHisAsnLleAspTyTLeuLysHis 131
Db 621 GCCC---AAAGAGAAAGACACCGACGCGCACAG-----CAACAGCA 659

QY 131 uSerLysValLysLysLleProSerAspPheGluLysSerGlyTyTThrAspLeuGlnG1 151
Db 660 GAGCAAA-----ACCAAGCAAAAGAGAGAGCGGCCCAACACCAACACAGC 704
QY 151 uArGlyAspAsnAspLleSerProPheSerGlyAspGlyGlnProPheLysAspLlePr 171
Db 705 AAAGGCAACCAACCATGGA-----AACAGCAGAC 737
QY 171 oGlyLysGlyGluAlaThrGlyProAspLeuGluLysAspLleGlnThrGlyPheAl 191
Db 738 GAACACCCGCAAGCAACGACGACGCAATACCAACAGCA-----777
QY 191 aGlyProSerGluAlaGluSerThrHisLeuAspThrLysLysProGlyTyTAsnGlu1 211
Db 778 ----CACGAAGAGAAAGCAGCAGCAGCAACCAAG-----807
QY 211 eProGluArGluGluAsnGlyLysAsnThrLleGlyThrArGAspGluThrAlaLysG1 231
Db 808 ----ACAGAGACAAAC-----ACGCAACACCAACAAAGACAGCA 842
QY 231 uAlaAspAlaValAspValSerLeuValGluLysSerAsnAspLleMetGlySerThrAs 251
Db 843 GCACACAGACACACAAAGACACATACCGGAATACAAAGC-----885
QY 251 nPheLysGluLeuProGlyArGlyLysArGlyValAspAlaGlySerGlnAsnAlaHi 271
Db 886 ----CGAGAGACACACAA-----ACGGAATACGACCAACACACA 920
QY 271 sGlnGlyLysValGluPheHisTyTProProAlaProSerLysGluLysArGlyGlu1 291
Db 921 TAAAGGACAAAC-----CGAGAGACACACAA-----962
QY 291 ySerSerAspAlaAlaGluSerThrAsnTyTAsnGluLleProLysAsnGlyLysGlySe 311
Db 963 CAGCCAAACAGAAAGAAACCAACCAAGACAGACACACCGCAAGACGACGAC 1022
QY 311 rThArGlyGlyValAspHisSerAsnArGAsnGlnAlaThrLeuAsnGluLysGlnAr 331
Db 1023 ACGAAACGAAGCA-----CACAAACAGAGAGACGAGGAGAAAGAAACCAACACAAAG 1076
QY 331 gPheProSerLysGlyLysSerGlnGlyLeuPro-----342
Db 1077 ACGAAAGAGAGCGAGCAACCAACCAACCAACCGGCGACCATACCGACACGCACTAAG 1136
QY 343 -----LleProSerArGlyLe 348
Db 1137 GCACACAAACAGCGCGAGAAACAAACAAAGAGAAACAGACAGACAGCAAGAGAGAG 1196
QY 348 uAspAsnGluLleLysAsnGluMetAspSerPheAsnGlyProSerHisGluAsnLle1 368
Db 1197 AGACCAACAGACGAGAGG-----ACCAAAAGCGGAAAGCGCAACACCAACAGACACA 1247
QY 368 eThHisGlyArGlyLysTyTThrTyTValProHisArGlnAsnAsnSer-----T 385
Db 1248 GCGAAACAGGAAGAAAG-----AACCAAAAGCAACGCAAGAAACAGAGA 1289
QY 385 hArGAsnLysGlyMetProGlnGlnLysGlySerTPrgLyArGlnProHisSerAsnA 405
Db 1290 CGAGCAACGAGAGAAACCA-----AGAAACACACACGCAAGAA 1328
QY 405 rGArGpHeSerSerArGArGArGAspAspSerSerGlu 417
Db 1329 GAAGAAACAGACAGAGACCGGACGACGCAAGCAGAGAC 1366

RESULT 10
BE907390 1337 bp mRNA linear EST 20-OCT-2000
LOCUS 601500123F1 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:3902020 5',
DEFINITION mRNA sequence.
ACCESSION BE907390
VERSION BE907390.1 GI:10400901
KEYWORDS EST.

ORGANISM	Strongylocentrotus purpuratus Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; Echinoidea; Euechinoidea; Echinoidea; Echinoidea; Strongylocentrotidae; Strongylocentrotus. 1 (bases 1 to 934) Cameron, R.A., Mallaras, G., Rast, J.P., Martinez, P., Biondi, T.R., Swartzell, S., Wallace, J.C., Pouška, A.J., Livingston, B.T., Wray, G.A., Eitelsohn, C.A., Lehrach, H., Britten, R.J., Davidson, E.H. and Hood, L.									
TITLE	A sea urchin genome project: Sequence scan, virtual map, and additional resources									
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)									
MEDLINE	20402566									
COMMENT	Contact: Cameron, RA, Davidson, EH, Hood, L Division of Biology 156-29 California Institute of Technology Pasadena California 91125, USA Tel: (626) 395-8421 Fax: (626) 793-3047 Email: acameron@caltech.edu Plate: 1006 row: N column: 12 Seq primer: 17 Class: BAC ends High quality sequence stop: 934.									
FEATURES	source 1..934 /organism="Strongylocentrotus purpuratus" /db_xref="taxon:7668" /clone="Plate=1006 Col=12 Row=N" /clone_lib="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library" /note="Organ: sperm; Vector: BAC3.6; BAC Clones in E-Coli DH10B"									
BASE COUNT	363 a 207 c 333 g 22 t 9 others									
ORIGIN										
Alignment Scores:										
Pred. No.:	0.000132 Length: 934									
Score:	150.50 Matches: 76									
Percent Similarity:	37.09% Conservative: 36									
Best local Similarity:	25.17% Mismatches: 131									
Query Match:	6.60% Indels: 60									
DB:	Gaps: 11									
US-09-700-696B-2 (1-430) x AZ186441 (1-934)										
QY	138	ProserasphegluglyserglytYrThrAspLeugluArglyAspAsnspIle	157							
DB	19	CCGAATCTTTAAAGTNNACCAACACCCCCCNAGAGAAAGAAACCAATTT	78							
QY	158	SerPropheserclYAspDlyglInProPhelyAspIlePro---	176							
DB	79	CAGGGGGGAGAAAGCCGANNCCCNAAAAAACCCGCCACGGGGGAACCCCGGAC	138							
QY	177	ThnglyProAspLeugluLyAspIlegInThrIlyPhealaglProsergluaIa	196							
	:::									
DB	139	AGGAACCCAGACGGAAGAAAGCAAGGCGACCGGAAACAAAGGCTT-TCAGAGAGC	197							
QY	197	GlUserThHisLeuAspThrlYslySPrOglYTYrAsnGlUleProslUArgglUglU	216							
	:::									
DB	198	AGC-----GAGAAAAAGAGACAGAAAGTGCAGCGCGACACAGAGAGAGAA	245							
QY	217	-----AsnGlYglYAsnThrlIeglYThrArgAspGlUThrAlaYsgluAlaAsp	233							
DB	246	CTTATCTCTTGGGGG-----GTGCNAAAGAGACAGACCCACAGAGAAAG	296							
QY	234	AlaValAspValSerLeuValglUglySerAsnAsp-----IleMetGlySer	249							
DB	297	GAGAGACCCCCCAGAGGGGAGAGGAGAAAGACCCCCCGGAAAGACAGAGAGAG	356							
QY	250	ThrAsnPhelYsglUleuProglYArAgglUglyAsnArgValAspAlaglYserGlInAsn	269							
DB	357	CAGACGCAAGAAAGACGAAAGCAAGAGAGAGAGCGA-----	395							

QY	270	AlahisglnlglylvsValglunphehisrtyrproprlaarproserlvsGltysArgrlys	289
Dd	396	-----CGGCACCAACAACACACACCAGCCGNAGGGGCCCGCGGGCCCCGGCCGACG	443
QY	290	GluglySerSerAspAlaAlagIuseThraSnTyraSngluileProLyAsnOlyLys	309
Dd	444	GAGACMAACAGACAAAGCGCAGAAAAGAACAGCACGAGAAAAAACCAGACAGCAC	503
QY	310	glySerThrArgLysGlyValAspHisSerAsnArgAsnGlnAlaThrLeuAsnGluLys	329
Dd	504	AAAGGAAAAACAGACAGAGAGAGAGAGAACAGCAAGCCAGACAGCGCAGCCAAAGCGAGGCG	563
QY	330	GlnArgPheProSerLysGlyLysSerGlnIleuProlleProSerArgGlyLeuasp	349
Dd	564	CggcGAGACAGACAAACACCCGAAAGCAGAGCGCCGACACACAGCGGAAGA-----	614
QY	350	AsngluileLysAsnGluMetAspSerPheAsnGlyProSerHisGluAsnIlellethr	369
Dd	615	--GAAGCGAAAGAGAGAGCGCAGCGAAAGCGGAAAGCAAGAAAGCAA-----	659
QY	370	HISglYArgLysTyrHisTyrValProHIsArGGLInAsnAsnSerThrArgAsnLysGly	389
Dd	660	-----CGAAGGAAAGGA	671
QY	390	-----MetProGlnGlyLysGlySerTrpgLYArgGlnProHisSerAsnArgarg	406
Dd	672	GCAGAAAGAGAGCCCGCAGAGAGAGAGAGAGAGCGGCGACACGCCG---AGCAACCCCGCC	728
QY	407	PheSerSerArgArgArgAspSerSer-----GluSerSerAspSer	421
Dd	729	GAAAGAAAGCAGAGAGCGCCGAGACAGACAGCAGAGAGAGAGAGAGAGGCCAAGAGCGC	788
QY	422	glySer 423	
Dd	789	GGGAGC 794	
RESULT 12			
LOCUS	BG287570	1366 bp	mRNA linear EST 21-FEB-2001
DEFINITION	602384569p1 NIH_MGC_93 Homo sapiens CDNA clone IMAGE:4513625 5',		
ACCESSION	BG287570		
VERSION	BG287570.1		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 1366)		
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/.		
COMMENT	Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: ATCC CDNA library Preparation: Life Technologies, Inc. DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov Plate: LHAM10400 row: 1 column: 18 High quality sequence stop: 192. Location/Qualifiers 1. 1366 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:4513625" /clone_1lb="NIH_MGC_93" /tissue_type="transitional cell papilloma, cell line" /lab_host="DH10B (phage-resistant)" /note="Organ: bladder; Vector: pCMV-SPORT6; Site_1: NotI;		
FEATURES			
source			

Site-2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.7 kb. Library enriched for full-length clones and constructed by Life Technologies.
Note: this is a NIH-MGC Library.
BASE COUNT 610 a 320 c 337 g 99 t
ORIGIN

Alignment Scores:

Pred. NO.: 0.000205 Length: 1366
Score: 150.50 Matches: 88
Percent Similarity: 34.45% Conservative: 66
Best Local Similarity: 19.69% Mismatches: 202
Query Match: 6.60% Indels: 91
DB: 12 Caps: 16

US-09-700-696b-2 (1-430) x BG287570 (1-1366)

```

QY      8 SeranlysgluasnThrHisAsnGlyLeuArgMetSerIleTyrProlySerThrGly 27
DB      211 AATTAACAGCAAAATAGAGAGAAACAAAGAAACAAAGAAACAAAGAAATACAGAGT 270
QY      28 AsnlysglyPhegluaSerGlyAspAspAlaIleSerIleuHisAspGlnGluTyr 47
DB      271 AACACAGGAGAGACAAACCGGACGACCGGCGCCATTAAACGCGAGAGACGTAG 330
QY      48 GylAlaAlaLeuIleArgAsnMetGlnHisIleMetGlyProValThrAlaIleLys 67
DB      331 AGCCCAAGAAAGAAAGAAAGACAGCAG-----CCGCAAGAAACGGTCAAA 375
QY      68 LeuLeuGlyGluGluAsnLysGluAsnThrProArgAsnValLeuAsnIleIleProAla 87
DB      376 CAACACTGGGAAAGCGCGGAAACCTCGAGACACACTG----- 417
QY      88 SerMetAsnTyrAlaLysAlaHisSerLysAspLysLysLysProGlnArgAspSerGln 107
DB      418 -----GGGGAACAACCCACCAACCCCAAGAGACAGCAG 453
QY      108 AlaGlnLysSerProValLysSerLysSerThrHisArgIleGlnHisAsnIleAspTyr 127
DB      454 ACACAAACAAAGCGCGGCAACGAGACCCCAACAAATATAG----- 492
QY      128 LeuLysHisLeuSerLysValLysLysIleProSerAspPhe-----GluGly 143
DB      493 -----ACAAATAATAAAAGATCCCGGAGAGGTATCAAGAACCCCAACAAAT 540
QY      144 SerGlyTyrThrAspLeuGlnGluArg-----Gly 153
DB      541 GTGGCAAAAGCGGACACACAGGCTCGCAACAGAGGTCAACACCGCCCAACAGAAC 600
QY      154 AspAsnAspIleSerProPheSerGlyAspGlyGlnProPheLysAspIleProGlyLys 173
DB      601 CCCAATGAA-----GCGCGTACGACAGACCATACAGACGCTCAAGGGAGG 648
QY      174 GlyAlaAlaThrGlyProAspLeuGlnGlyLysAspIleGlnThrGlyPheAlaGlyPro 193
DB      649 AACCAATGGGGCACACACAGACAAACAAACAAACCAACCCCAAGAGAGACATGGGGC 708
QY      194 SerGlnAlaGluSerThrHisLeuAspThrLysLysProGlyTyrAsnGluIleProGlu 213
DB      709 AACCAAAACGAGAAAA--CACACGAGAGAAAAAGCAAGAACCAAAAGAGACCAAAAC 765
QY      214 ArgGluGluAsnGlyLysAsnThrIleGlyTyr----- 224
DB      766 AAGGAAGAGAGAAACAGACAGAACG--GGAACACACACCAACAGAGAGAGAGAGACA 822
QY      225 -----ArgAspGluThrAlaLysGluAlaLysAlaValaLysValSerLeu 239
DB      823 GAACACAGAGAGAAAGAAAGAAAGAGAGAAAGAGAAACCAAGACCAAGACCGCG-- 879
QY      240 ValGluGlySerAsnAspIleMetGlySerThrAsnPheLysGluLeuProGlyArgGlu 259
DB      880 ----GAGAGCGGTAAAGCAACACCCCGAGCGAGCAAGAGAGAGAGAA-----CGAAG 927

```

```

QY      260 GlyAsnArgValAspAlaGlySerGlnAsnAlaHisGlnGlyLysValGluPheHisTyr 279
DB      928 AACACCCAGCCACACACACCGCGCGCTCGATACACACCCAGAGCCCGCA----- 978
QY      280 ProProAlaProSerLysGluLysArgLysGluGlySerSerAspAlaIleGluSerThr 299
DB      979 CCGACGACGCGCAAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1038
QY      300 AsnTyrAsnGluIleProLysAsnGlyLysGlySerThrArgLysGlyValAspHisSer 319
DB      1039 AAG-----CAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1083
QY      320 AsnArgAsnGlnAlaThrLeuAsnGluLysGlnArgPheProSerLysGlyLysSerGln 339
DB      1084 AAGCGACACCGCGCAAGC-----CCGCAACGACGCGCGAGAGAGAGAGAGAGAGAGAG 1134
QY      340 GlyLeuProIleProSerArgLysLeuAspAsnGluIleLysAsnGluMetAspSerPhe 359
DB      1135 CACACAAAGATGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1170
QY      360 AsnGlyProSerHisGluAsnIleIleThrHisGlyArgLysTyrHisTyrValProHis 379
DB      1171 ACAGACCGACGACAGACAAACACAGCCGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1221
QY      380 ArgGlnAsnAsnSerThrArgAsnLysGlyMetProGlnGlyLysGlySerTrpGlyArg 399
DB      1222 AAGCGGTCAAAACAAACCAACAGACAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1281
QY      400 GlnProHisSerAsnArgArgPheSerSerArgArgArgAspAspSerSerGluSerSer 419
DB      1282 GAACCGCGGCAAGCGCGGTGAGACCCACACAGCGCGCGCGAGCAACAGCAGCGCAGACA 1341
QY      420 AspSerGlySerSerSerGlu 426
DB      1342 GCAAAACGACGCGCCAGACACAA 1362

```

RESULT 13
BE779053
LOCUS
DEFINITION
601464816f1 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:3868275 5',
mRNA sequence.
ACCESSION
BE779053
VERSION
BE779053.1 GI:10200251
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1367)
NIH-MGC <http://imgc.ncbi.nlm.nih.gov/>.
NATIONAL INSTITUTES OF HEALTH, MAMMALIAN GENE COLLECTION (MGC)
Unpublished (1999)
CONTACT: Robert Strausberg, Ph.D.
Email: cga@dbi-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
Clone distribution: MGC clone distribution information can be
[found through the I.M.A.G.E. Consortium/LNL at:](http://imgc.lnl.gov)
<http://imgc.lnl.gov>
Plate: LLM9615 row: p column: 04
High quality sequence stop: 108.
Location/Qualifiers
1. 1367
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3868275"
/clone_id="NIH_MGC_67"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.

FEATURES

source

BASE COUNT 363 a 186 c 456 g 63 t
ORIGIN

Alignment Scores:

Pred. No.: 0.000194 Length: 1068
Score: 149.50 Matches: 81
Percent Similarity: 41.03% Conservative: 47
Best Local Similarity: 25.96% Mismatches: 140
Query Match: 6.56% Indels: 44
DB: 12 Gaps: 12

US-09-700-696b-2 (1-430) x BG538300 (1-1068)

```

QY 131 LeuSerLysValLysLysIleProSerAspPheGluGlySerGlyThrAspLeuGln 150
Db 112 ATAAAAAACCCCGCGGCAACCGAGGTCTGACAAATCCAGATGAGTCTGAGGTC 171
QY 151 GluArgGlyAspAsnAspIleSerProPheSerGlyAspGlyGlnProPheLysAspIle 170
Db 172 ATTACTGATCTATCAACAGAGATCCA-----AGCGAGCTCGAAGATTCGAAAATTTA 225
QY 171 ProGlyLysGlyGluAlaThrGlyProAspLeuGluGly-LysAspIleGlnThrGlyPh 190
Db 226 AGCGGCCCCACGAG-----CCCTGAGACAAACAAACAGACGCAACGGGCGAC 273
QY 190 eAlaGly---Pro-SerGluAlaGluSerThrHisLeuAspThrLysLysProGlyTyrA 209
Db 274 CGCAGGAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 333
QY 209 sngluIleProGluArgGluGluGlnGlyGlnGlyGlnGlyGlnGlyGlnGlyGlnGly 229
Db 334 ACAGAGAGACACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 393
QY 229 IalysGlu-----AlaAspAlaValAspValSerLeuValGluGlySerAsnAspIleW 247
Db 394 GAAAGAGGACACCCCGCACAGGCGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 453
QY 247 etGlySerThrAsnPhe-LysGluLeuProGlyArg----- 258
Db 454 GAGGAGATCGGGCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 513
QY 259 GluGlyAsnArgValAspAlaGlySerGlnAsnAlaHisGlnGlyLysValGluPheHis 278
Db 514 GAAGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 564
QY 279 TyrProAlaProSerLysGluLysArgLysGluGlySerSerAspAlaAlaGluSer 298
Db 565 -----GCGGCGACAAAAGCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
QY 299 ThrAsnTyrAsnGluIleProLysAsnGlyLysGlySerThrArgLysGlyValAspHis 318
Db 601 ACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
QY 319 SerAsnArgAsnGlnAlaThrLeuAsnGluLysGlnArgPheProSerLysGlyLysSer 338
Db 661 GGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 702
QY 339 GlnGlyLeuProIleProSerArgGlyLeuAspAsnGluIleLysAsnGluMetAspSer 358
Db 703 GAAGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 759
QY 359 PheAsnGlyProSerHisGluAsnIleLeuThrHisGlyArgLysTyrHisTyrValPro 378
Db 760 AGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 807
QY 379 HisArgGlnAsnAsnSerThrArgAsnLysGlyMetProGlnGlyLysGlySerTProGly 398
Db 808 GGGAGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 867
QY 399 ArgGlnProHisSerAsnArgPheSerSerArgArgAspAspSerSerGluSer 418
Db 868 AGAAGAGGGGCGCAACGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 921

```

QY 419 SerAspSerGlySerSerSerGluSerAsp 428
Db 922 GACGACAAAGGAG 951

Search completed: April 17, 2003, 02:06:15
Job time : 1944 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: April 16, 2003, 19:48:00 ; Search time 50 Seconds

(without alignments)
2637,420 Million cell updates/sec

Title: US-09-700-696B-2

Perfect score: 2279
Sequence: 1 VKKEYSISNKNENTHNGLRMS.....RRDSSSSSDSGSSSDG 430

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-O=cg2_1/USFTO.spool/US09700696/runat_11042003_101726_12446/app.query.fasta_1.583
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPT=0
-LISTEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blsum62 -TRANS=human40.cdi
-LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USRR=US09700696 -CGN 1.1.40 -runat_11042003_101726_12446 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Issued_Patents_NA:*

1: /cg2_6/ptodata/1/lna/5A.COMB.seq:*
2: /cg2_6/ptodata/1/lna/5B.COMB.seq:*
3: /cg2_6/ptodata/1/lna/6A.COMB.seq:*
4: /cg2_6/ptodata/1/lna/6B.COMB.seq:*
5: /cg2_6/ptodata/1/lna/PCUS.COMB.seq:*
6: /cg2_6/ptodata/1/lna/backfilest1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	150	6.6	6755	3	US-08-931-999-4
2	140	6.1	2690	1	US-08-524-757-11
3	138.5	6.1	2384	1	US-07-814-964-10
4	138.5	6.1	2384	1	US-08-258-442-10
5	138.5	6.1	2384	1	US-08-328-809-5
6	137	6.0	9370	1	US-08-320-559-27
7	137	6.0	9370	1	US-08-320-559-27
8	137	6.0	9370	1	US-08-320-559-27
9	137	6.0	9370	1	US-08-320-559-27
10	137	6.0	9391	3	US-08-545-860D-25
11	137	6.0	9391	3	US-08-545-860D-25
12	137	6.0	9391	5	PCT-US94-04496-25

13	134.5	5.9	4084	2	US-08-568-459A-1	Sequence 1, Appl1
14	134.5	5.9	4084	2	US-08-487-826B-1	Sequence 1, Appl1
15	134.5	5.9	4084	4	US-09-210-288-1	Sequence 1, Appl1
16	134.5	5.9	4084	6	5198347-5	Sequence 1, Appl1
17	129	5.7	3157	6	5198347-3	Sequence 1, Appl1
18	127.5	5.6	3707	1	US-08-118-101A-1	Sequence 1, Appl1
19	126.5	5.6	1393	5	PCT-US93-07261-12	Sequence 1, Appl1
20	126.5	5.6	3773	4	US-09-130-242-1	Sequence 1, Appl1
21	125.5	5.5	3825	4	US-09-208-742-3	Sequence 1, Appl1
22	125.5	5.5	5173	4	US-08-801-308-2	Sequence 1, Appl1
23	124.5	5.5	3552	4	US-09-134-001C-693	Sequence 1, Appl1
24	123	5.4	1503	4	US-08-999-774A-11	Sequence 1, Appl1
25	123	5.4	3534	4	US-09-134-001C-2269	Sequence 1, Appl1
26	123	5.4	4766	5	PCT-US93-07261-10	Sequence 1, Appl1
27	122	5.4	2073	4	US-09-134-001C-1731	Sequence 1, Appl1
28	122	5.4	43280	2	US-08-804-227C-1	Sequence 1, Appl1
29	121	5.3	2004	1	US-08-471-033-18	Sequence 1, Appl1
30	121	5.3	2004	2	US-08-471-033-18	Sequence 1, Appl1
31	121	5.3	2004	2	US-08-463-483A-18	Sequence 1, Appl1
32	121	5.3	2004	2	US-08-471-046A-18	Sequence 1, Appl1
33	121	5.3	2004	2	US-08-471-046A-18	Sequence 1, Appl1
34	121	5.3	2004	2	US-08-470-566B-18	Sequence 1, Appl1
35	121	5.3	2004	3	US-08-469-334-18	Sequence 1, Appl1
36	121	5.3	2655	1	US-08-471-033-17	Sequence 1, Appl1
37	121	5.3	2655	2	US-08-463-483A-17	Sequence 1, Appl1
38	121	5.3	2655	2	US-08-471-046A-17	Sequence 1, Appl1
39	121	5.3	2655	2	US-08-470-566B-17	Sequence 1, Appl1
40	121	5.3	2655	2	US-08-469-334-17	Sequence 1, Appl1
41	121	5.3	2655	2	US-08-469-334-17	Sequence 1, Appl1
42	121	5.3	2655	2	US-08-469-334-17	Sequence 1, Appl1
43	120.5	5.3	4507	2	US-08-568-459A-3	Sequence 1, Appl1
44	120.5	5.3	4507	2	US-08-487-826B-3	Sequence 1, Appl1
45	120.5	5.3	4507	4	US-09-210-288-3	Sequence 1, Appl1

ALIGNMENTS

RESULT 1
US-08-931-999-4
Sequence 4, Application US/08931999
Patent No. 6043219

GENERAL INFORMATION:
APPLICANT: Iandolo, John J.
APPLICANT: Crupper, Scott S.
TITLE OF INVENTION: Broad Spectrum Chemotherapeutic Peptide
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: Hovey, Williams, Timmons & Collins
STREET: 2405 Grand Boulevard, Suite 400
CITY: Kansas City
STATE: Missouri
COUNTRY: U.S.A.
ZIP: 64108

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/931,999
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/710,561
FILING DATE: 19-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Collins, John M.
REGISTRATION NUMBER: 26,262
REFERENCE/DOCKET NUMBER: 25043-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 816/474-9050
TELEFAX: 816/474-9057
INFORMATION FOR SEQ ID NO: 4:

Oy	2	AsnysgluTyrSerIleSerAsnLysGluAsnThrHisAsnGlyLeuArgMetSerIle	21
	111	111
Dd	4164	AACCAAAACCAACGAGAGGAGCAACCAACCCACACAGACAGAAAGCAA	4217
Oy	22	TyrProLysSerThrGlyAsnLysGlyPheGluAspGlyAspAspAlaIleSerLysLeu	41
	111	111 111	111
Dd	4218	-----AAGCAAAACGGCGCGAAA-----AACGACCAAAAAAAACAACAAAACA	4259
Oy	42	HisAspGlnGlu-----	45
	111	
Dd	4260	CACAAGACAGAGAAAAAGCCACCGACACAAAACCCGAGCAAAAACCCAGAAAAACAAA	4319
Oy	46	---GluTyrGlyAlaIleAlaLeuIleArgAsnAspMetGlnHisIleMetGlyProValThr	64
	111	111
Dd	4320	GGAAAAAAGCGAAAAAAGACACACCAAAAGCAAAAAGAAAGAAAGAAAGCGCGAAAAA	4379
Oy	65	AlaIleLysLeuLeuGlyGluGluAsnLysGluAsnThrProArgAsnValLeuAsnIle	84
	111	111
Dd	4380	AGCAAGAAAA-----AAAGCAAAACACAGAACCCACCAACAAACGACAAAAAACA	4433
Oy	85	IleProAlaSerMetAsnTyrAlaIleLysAlaHisSerLysAspLysLysLysProGlnArg	104
	111	111
Dd	4434	CAAGACGACAGAAAAAGCCAGAACCCAGAGAAAAAAGGAAAAAACCAACAAAAAAG	4493
Oy	105	AspSerGlnAlaGln-----Lys	110
	111	111
Dd	4494	GAAACGAAAAACCAAAAGCAGAAAGCAACAAACAAAGAGCGGAGACGAGAGAAAAA	4553
Oy	111	SerProValLys-----SerLysSerThrHisArgIleGlnHisAsn	124
	111	111
Dd	4554	ACACCCCGGCAAAAAACAAACAAAAAGACACACCAACCAACCCCAAGAGACAGAAAGAAAC	4613
Oy	125	IleAspTyrLeuLysHisLeuSerLysValLysLysIleProSerAspPheGluLysSer	144
	111	111111	111111
Dd	4614	-----ACAACGAAAAA-----GAAACGCGGAGAGAAAGA	4643
Oy	145	GlyTyrThrAspLeuGlnGluArgGlyAspAsnAspIleSerProPheSerGlyAspGly	164
	111	111
Dd	4644	GGAAACAGCCAGAAACAGACAAACCAACAGAACCGGAGGAGACCCAGAACAAAAAGAAAGC	4703
Oy	165	GlnProPheLysAspIleProGlyLysGlyGlyAlaThrGlyProAspLeuGlnGlyLys	184
	111	111 111	111 111
Dd	4704	-----AAAACGAGAGAGGACAAACCGCAAAACAGGAACACAGAACACAAAGAAAA	4755
Oy	185	AspIleGln-----	187
	111	
Dd	4755	GAAAAACACAGAAACCGAAACGGAACACAGACAGAAACAAAAACGAAACGACAAAGA	4814
Oy	188	-----ThrGlyPheAlaGlyProSerGlyAlaGlySerThrHis	200
	111	111111 111	111 111

Db	4815	AAAGCGCGCGACGACAAAAAAGAGAG -AGAGACCCCGAGAGAGACACCCCGAC	4873
Qy	201	LeuaspThrLysLys-----	205
Db	4874	ACCCAGAAAAAATACAGACAAAGAGAGCGAAAAACACAGAAAAAAGAAACCGACGCG	4933
Qy	206	-----ProGlyTyrAsnGluTLeProGluTArgGlu	215
Db	4934	AAAGAGAGCGAAAAACGAAAAAAGAAAAAAGAAACACAAAAAAGGAGAGAAAAAAG	4993
Qy	216	GluAsnGlyGlyAsnThrTLeGlyThrArgAspGluThrAlaLysGluAlaAspAlaVal	235
Db	4994	CACACACGGGGGGCGCAACAGCGAGCACGACGACAGCAAGAAAAACGGGAA-----	5044
Qy	236	AspValSerLeuValGluGlySerAsnAspTLeMetGlySerThrAsnPhenGlyLeu	255
Db	5045	-----AAACAAAAAACGACAAAAAAGCAAAAGGCAAAACCCCAAAAGAACCA	5083
Qy	256	ProGlyArgGluGlyAsnArgValAspAlaGlySerGlnAsnAlaHisGlnGlyLysVal	275
Db	5084	AAAAACAGCGCAGCG-----CAAGCGCG-----AACACACACAAACAAACAAAG	5128
Qy	276	GluPheHisThrProAlaProSerLysGluLysArgLysGluGlySerSerAspAla	295
Db	5129	ACAGGAGACACAAAAACCGCGCAAAAGGGGAAACAAAAAGGAGGAAAGGGGGGAAA-----	5182
Qy	296	AlaGluSerThrAsnTyrAsnGluTLeProLysAsnGlyLysGlySerThrArgLysGly	315
Db	5183	-----AAAGAAAAAGGCAACAGGGGAGGGGCGCAAGAGAAACGCGCA	5224
Qy	316	ValAspHisSerAsnArgAsnGlnAlaThrLeuAsnGluLysGlnArgPheProSerLys	335
Db	5225	AAAAAAGAAAAAACCGGACCCAGAAACGAAACGAAACCAAGAAAAAAGAAAGAGCAAAAG	5284
Qy	336	-----GlyLysSerGlnGlyLeuProLleProSerArgGlyLeuAspAsnGluLys	353
Db	5285	AAAAAGGGCAAAAGCAGA-----	5302
Qy	354	AsnGluMetAspSerPheAsnGlyProSerHisGluAsnTLeIleThrHisGlyLys	373
Db	5303	-----AAAGAACACACCAACAAAGGGGAGCAAGGACACAC-----AAAAACCGGAAA	5350
Qy	374	TyrHisTyrValProHisArg-----GlnAsnAsnSerThr	385
Db	5351	AGAGCGGAAAAACCAAAACGAAACACCCGAAAGCAAGAGCAAGCAACGAAACGAAAG	5410
Qy	386	ArgAsnLysGlyMetProGlnGlyLysGlySerTyrGlnArgGlnProHisSerAsnArg	405
Db	5411	AGACACAGCGGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAACCAACCGCGGAAAGCCACAGA	5470
Qy	406	ArgPheSerSerArgArgArgAspAspSerSerGluSerSerAsp-----SerGlySer	424
Db	5471	AGAAAGGCAACGAGAGGAGGAGAACAGAACCAACGACGACAGAGAAAGAAAGAACGGGA	5530
Qy	425	SerGluSerAspGly	429
Db	5531	ACCCACACAGAGGA	5545
RESULT 2			
US-08-524-757-11			
Sequence 11, Application US/08524757			
Patent No. 5792634			
GENERAL INFORMATION:			
APPLICANT: Conway, Ronald C.			
APPLICANT: Conway, Joan W.			
APPLICANT: Bradsher, John N.			
TITLE OF INVENTION: RNA Polymerase Transcription Factor			
NUMBER OF SEQUENCES: 37			
CORRESPONDENCE ADDRESS:			
ADDRESSEE: RICHARDS, MEDLOCK & ANDREWS			
STREET: 1201 Elm Street, Suite 4500			
CITY: Dallas			
STATE: TX			

```

COUNTRY: US
ZIP: 75270-2197
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/524,757
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/13621
FILING DATE: 29-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/160087
FILING DATE: 30-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Harre, John A.
REGISTRATION NUMBER: 37,345
REFERENCE/DOCKET NUMBER: B35006CIPICP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (214) 939-4500
TELEFAX: (214) 939-4600
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 2690 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 33..2351
US-08-524-757-11

Alignment Scores:
Pred. No.: 0.000293
Score: 140.00
Percent Similarity: 32.08%
Best Local Similarity: 20.52%
Query Match: 6.14%
DB: 1
Caps: 17

US-09-700-696b-2 (1-430) x US-08-524-757-11 (1-2690)
QY 23 ProlysserThrglyasnlysglyphegluasplasyaspsalalelserylsleuHis 42
Db CCTGGAAAGGCGACAGCAATGCTTCAGGACAGACTCGGGCCACCAAGACGACAC 719
QY 43 AsplngluluglyrglyalaalaaleuileargasnmetelglnHisilemetglyPro 62
Db CTGGGTGAACCCCATGGGAAAGGGGTGTGACTCAAAACAAGAGGAC 767
QY 63 ValhralailelylsleuGlugluasnlysgluasnhrproargasnValleu 82
Db 767 ----- 767
QY 83 AsnlelleproalaaserMetasnTYrAlaLysAlaHisSerlyAspLysLysPro 102
Db 768 ----- AAATCTCCCAACAAGCAAAAGCCCGCTG 797
QY 103 GluArgaspserglnalaglnlyserprovallysserlyserthrhrisatgllgln 122
Db 798 GATGCCAAGATGATGAAAGCCCTCTGTGTGACAGCAAGAAATCAAC 848
QY 123 HisasnleasptYrleuYshlsleuSerlySvallyslleproserAspPheglu 142
Db 849 ----- AAGGCCCTCTCCAA 863
QY 143 GlyserglytyrthrAspPhegluIngluargglyAspAsnAspLleserProPheSergly 162
Db 864 ----- GAGGAGAACCGCAAGGCCACCTCAGG 890

```

```

QY 163 AspGlyGlnProPheLysAspLleProGlyLysGlyGluAlaThrnglyProAspLeuGlu 182
Db 891 GACATGCAAGGAGAAACCG---CCCTCTGTGGCGTAAAGAAAGAAAGACAGAGAG 947
QY 183 GlyLysAspLleGlnThrGlyPheAlaGlyProSergLysLysLysThrHisLeu 201
Db 948 GGCAGCACCTGAAAGAAAGTGTGGCTCCCTCAGAGGCGCGCTCAGCAACCCACTG 1007
QY 202 AspThrLysLysProGlyTYrAsnGluileProGluArgGluGluasnGlyGlyAsnThr 221
Db 1008 -----AAAAGCCAAAGCACAGAGAC---CCAGAGAAACCCAA 1043
QY 222 IleGlyThrArgAspGluThrAlaLysGluAlaAspAlaValAspValSerLeuValGlu 241
Db 1044 -----TTGACAAAGAACGAGTGTGCTGACAGCTTTGACACAGA-----AAA 1088
QY 242 GlySerAsnAspLleMet-----GlyserThrAsnPhelysgluLeu 255
Db 1089 GGACAGAGAGACCTGTGGCCCAAGGTAAAGAGAGGTTCTTAACAACCTAAAGACTCCA 1148
QY 256 ProGlyArgGluGlyAsnArgValAspAlaGlySergLysAsnAlaHisGlnGlySval 275
Db 1149 GAAGGAAAGTCAAAACTAATTGTGATGAGAAAGTCACTGGGCTCC---CTCCCTAAAGTT 1205
QY 276 -----GluPheHisTYrPro----- 280
Db 1206 GAGGAGACAGATATGAGATGATTCAGACACCAACCATGCTTTGAATCTACCTC 1265
QY 281 -----ProAlaProSerysglyLysArgGlyGlyLysSerAspAlaAlaGlnSer 298
Db 1266 AGCTATGACACGCCCGGAGAAAGAAAGAAAGATTTGAAACT---TCAGCCAGCGCA 1322
QY 299 ThrAsnTYrAsnGluileProLysAsnGlyLysGlySerThrArgGlySvalAspHis 318
Db 1323 CTGGAGATTAAGACTTAAATAAATGACTCTAAAGACACTGTGTAACCTTGCACTCA 1382
QY 319 SerAsnArgAsnGlnAlaThrLeuasnGlyLysGlnArgPheProSerysglyLysSer 338
Db 1383 GTTCAGAAATTCACCAAGGTGAACCAAAACCAAGTCAAGAAACCGCTGAGCTGTTTA 1442
QY 339 GlnGlyLeu----- 341
Db 1443 GCCAAGCTGAGAAAGTGCCTGATGTGTTGCCAGTGTGCCAGACCTCCGTTACCCGCG 1502
QY 342 -----ProIleProSerArgGlyLeuAspAsnGluileLysasnGlu 355
Db 1503 ATACAGCCCAATTACCTGCTCAGCTTCCTTCGAGGTGATTCCTCTCCAGCCCAAG 1562
QY 356 MetAspSerPheAsnGlyProSerHisGluasn----- 366
Db 1563 CCAAAAGCTTCTCTACACCCAGAGAAAGAAAGCTGATTTACTGGCGGCAGAAATG 1622
QY 367 -----IleIleThrHisGlyArgLysTYrHisTYrValPro----- 378
Db 1623 AATTCAGATGACAGTGTATCTGTTCACAAAGTGTGCTATCTCTTAATAATGATGACC 1682
QY 379 -----HisArgGln 381
Db 1683 TTGCACACGCAA 1694

RESULT 3
US-07-814-964-10
; Sequence 10, Application US/07814964
; Patent No. 5359047
GENERAL INFORMATION:
APPLICANT: Donahue, Brian A.
APPLICANT: Toney, Jeffrey H.
APPLICANT: Bruhn, Suzanne L.
APPLICANT: Pili, Pieter M.
APPLICANT: Brown, Steven
APPLICANT: Kelleet, Pauli
APPLICANT: Essigmann, John M.

```



```

Oy      413 pspserSerIuserraspserglYserSerserJlu----- 426
          |||| | ||| |||||||
Db      2117 GCAGACATCCACCAGCTTCGACGACGAGNAGAACAGAGCCTGCCAAGAAAGAGCA 2176
Oy      427 -----SeraspJlasp 430
          |||||
Db      2177 GCCCATCGACGCGCGAT 2195

RESULT 4
US-08-258-442-10
; Sequence 10, Application US/08258442
; Patent No. 5670621
; GENERAL INFORMATION:
; APPLICANT: Donahue, Brian A.
; APPLICANT: Toney, Jeffrey H.
; APPLICANT: Bruhn, Suzanne L.
; APPLICANT: Pil, Pieter M.
; APPLICANT: Brown, Steven
; APPLICANT: Kellelt, Patti
; APPLICANT: Essigmann, John M.
; APPLICANT: Lippard, Stephen J.
; TITLE OF INVENTION: DNA Structure Specific Recognition
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: 2 Millitia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/258,442
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/539,906
FILING DATE: 18-JUN-1990
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: MTW-4787AAA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2384 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Drosophila melanogaster
IMMEDIATE SOURCE:
CLONE: Drosophila SSNP - composite sequence
POSITION IN GENOME:
CHROMOSOME/SEGMENT: 2
MAP POSITION: 60A 1-4
FEATURE:
NAME/KEY: CDS
LOCATION: 123..2291
US-08-258-442-10

Alignment Scores:
Pctd. No.: 0.000345 Length: 2384
Score: 138.50 Matches: 99

```

Percent Similarity:	31.91%	Conservative:	50
Best Local Similarity:	21.20%	Mismatches:	136
Query Match:	6.08%	Indels:	182
DB:	1	Gaps:	23
US-09-700-696B-2 (1-430) x US-08-258-442-10 (1-2384)			
QY 16 GILEUATGMSerIeTYrProLySserThGly-ASrLYGGLyPhGluASpLYs	35		
DB 1187 GAGCGAGGATTCATCTATATCCCAAGACCGCGTCGATATCCGGCTTGGAGG----	124		
QY 35 PASPALIeSerLYsLeuNHISaRgInGInGluTYrGLyALA-----	49		
DB 1242 -----ATTAGTTCCTGAACCTTGGCCCGACGGCGGATCCACCGCATCTTGCATT	129		
QY 50 -----ALaLeuILeArGAsnMetGlnHISileMeGLyProVALThALaILeLYsLe	68		
DB 1295 CGAAGTCAGCTCAAGAACGAACTGTGTACATCTTCTCCATC-----	1340		
QY 68 uLEuGLyGLuSAsnLYSGluSAnThPrOArGAsnVALeASnILleProLaSe	88		
DB 1341 -----GAGAGAGAGAG-----	135		
QY 88 rMetAsnTYrALySALaNHISerLYsAsPLySLySProGInArGAsPserGInAl	108		
DB 1353 -----TATCCAAAGCTC-----	136		
QY 108 aGInLYSerProVALySserLYsSerThrHISArGLIeGInHISAnILeAsPTyLe	128		
DB 1355 -----	137		
QY 128 uLYs-----HISLeuSerLYsVALySLySILeProSerAsPheGLuGLySe	144		
DB 1376 CACACAGAGAAgTTCATGTCTCAGCAACATGGCGAAG-----GACAAAG	1422		
QY 144 rGLYTYrThrAsPLeuGLInGLuArGGLyAsPAsnAsPILeSerProPheSerGLyAsPcI	164		
DB 1421 CGGTCCTAAAGAGCTG-----GACTTGTGTATTCGGAC-----AACGAGNA	1467		
QY 164 yGLnProPheLYsAsPILeProGLy---LYSGInGLyNALHnGLyProAsPLeuGLI	183		
DB 1463 CGAAGCAGATGCCATGTGGCTGGCTCGAAGCTGAGCGAGGAAGAAAGAGAGAGCA	1522		
QY 183 YLYsAsPILeGLInThrGLyPheALaGLyProSerGLuALGLuSerThrHISLeuAsPTh	203		
DB 1523 CGAGAG-----GGCAGCTGGATGAAGAGTCCAGCAGATGAGAGACTT	1564		
QY 203 rLYsLYSProGLYrThrAsnGLuILeProGLuArGGLuSAsnGLyGLyAsnThILeGL	223		
DB 1565 C-----AAAGCCACAGAGAAAGAGAGTGGCCGAGAGAGTGAACAGCAGATGTGAGAG	1621		
QY 223 YThrArGAsPGLuThrALaLYSGInuALAsPArALaSPArLeuVALGLuGLySe	243		
DB 1622 TGATTCGAGCAT-----GACAGCGATGCTAGTGSC-----GGCGG	1657		
QY 243 rASnAsPILeMeGLYSerThrAsnPhrLYSGluLeuProGLuArGGLuGLyAsnArGyA	263		
DB 1658 AGCGAGACAGCGAGCGGCCAAGAAAAAGAGAGAAAGTCCGAAAGAAAGACAAAAA	1717		
QY 263 lAsPArAGLYSerGLYAsnALaNHISGLInGLyLYsVALGLuPhnHISrTYrProGLAr	283		
DB 1718 GAGAAAAAACACAAAGAGAGAGAGAGACAAAG-----AAACG	1756		
QY 283 oSerLYSGInLYArGLyGLuGLySer-----	292		
DB 1757 CTCCTAAGAAAGAAAGAGACTTGGCAAAACCAAGCGCGCACACCGCTTTCATCTCTG	1816		
QY 293 -----SerAsPArALaAGLuser-----ThrAsnTYrAsnGLuIL	304		
DB 1817 GCTGAACGACACCGCGAGAGACATCAAGAGGGAANAATCCGGCGATTAAGAGTTACCGAGAT	1876		
QY 304 eProLYSAsnGLyLYSGLYSerThrArGLyGLyVALAsPILSerAsnArGAsnGLnAl	324		

Db 1877 CGCCAGAGAGGGCGGCGAGATGTGGAAGAGCTGAAGAGACAGTCCAGATGGAGAGATGC 1936
 QY 324 arhrleuasnglulysglinarphproserlysglylyserglnglyleuProIlepr 344
 Db 1937 GGGGCGCAAGACAGACAGCGCTAC----- 1961
 QY 344 oSerArGlyLeuAspnsnluilelysAsnnglumetAspSerPheAsnnglyProserH1 364
 Db 1961 ----- 1961
 QY 364 sgluasnllellehrhslsclayrlystyryhstYrValProHisArglnAsnAsnSe 384
 Db 1962 -----CACAGCGAGATCGCGAAGCTTACAGCGGCGGTGACAG 2005
 QY 384 rThrArGAsnLysGlymetProGlnGlyLysGlySerTyrPglYArg-----GlnProH1 402
 Db 2006 CGACACAGACAGAGGCT-----GGAAAGTCTCTCCAGAAAGCGCAAGAGCGCTTC 2056
 QY 402 sSerAsnArGArg-----PheSerSerArGArg-----ArgAs 413
 Db 2057 TCCATCCAGAAAGCGCAATACCTCGGCGAGCGCTTCAAGAGCAAGAGATACATTTCGA 2116
 QY 413 pAspSerSerGluSerSerAspSerGlySerSerSerGlu----- 426
 Db 2117 CGAGGACTCCAGCCAGCTCCGACGACGAGAAAGACACAGAGCTTCGCAAGAAAGACAA 2176
 QY 427 -----SerAspGlyasp 430
 Db 2177 GCCCCCATCCGACGCGCAT 2195

RESULT 5

US-08-328-809-5
 Sequence 5. Application US/08328809
 Patent No. 5705334

GENERAL INFORMATION:

APPLICANT: Lippard, Stephen J.
 APPLICANT: Essigmann, John M.
 APPLICANT: Donahue, Brian A.
 APPLICANT: Toney, Jeffrey H.
 APPLICANT: Bruhn, Suzanne L.
 APPLICANT: Pili, Pieter M.
 APPLICANT: Brown, Steven
 APPLICANT: Kellelt, Patli
 TITLE OF INVENTION: Uses For DNA Structure-Specific
 NUMBER OF SEQUENCES: 8
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Patent Administrator, Testa, Hurwitz & Thibault
 STREET: 53 State Street
 CITY: Boston
 STATE: MA
 COUNTRY: USA
 ZIP: 02109
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/328, 809
 FILING DATE:
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Fenton, Gillian M.
 REGISTRATION NUMBER: 36,508
 REFERENCE/DOCKET NUMBER: MIT-023 (5473/24)
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617-248-7000
 TELEFAX: 617-248-7100
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2384 base pairs
 TYPE: nucleic acid

STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 ORIGINAL SOURCE:
 ORGANISM: Drosophila melanogaster
 IMMEDIATE SOURCE:
 CLONE: Drosophila SSRP - composite sequence
 POSITION IN GENOME:
 CHROMOSOME/SEGMENT: 2
 MAP POSITION: 60A 1-4
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 123..2291
 US-08-328-809-5
 Alignment Scores:
 Pred. No.: 0.000345 Length: 2384
 Score: 138.50 Matches: 99
 Percent Similarity: 31.91% Conservative: 50
 Best Local Similarity: 21.20% Mismatches: 136
 Query Match: 6.08% Indels: 182
 DB: Gaps: 23

US-09-700-696B-2 (1-430) x US-08-328-809-5 (1-2384)

QY 16 GlyLeuArGmetSerIleTyrProLysSerThrGly-AsnLysGlyPheGluAspGlyAs 35
 Db 1187 GGAGCGAGGATTCATCTATTCACAGCACCGCTGCATATCCGCTTGAGAG----- 1241
 QY 35 pAspAlaIleSerLysLeuHisAspGlnGluGlnTyrGlyAla----- 49
 Db 1242 -----ATTAGTCTCTGAACTTGGCCGAGCGCGGATCCAGCGACTTTCGACTT 1294
 QY 50 ----AlaLeuIleArgAsnAsnMetGlnHisIleMetGlyProValThrAlaIleLysLe 68
 Db 1295 CGAAGTGACGCTCAAGACGGAACGTTCACATCTCTCCATC----- 1340
 QY 68 uLeuGlyGlnGluAsnLysGluAsnThrProAlGAsnValLeuAsnIleIleProAlase 88
 Db 1341 -----GAGAAAGAGAG----- 1352
 QY 88 rMetAsnTyrAlaLysAlaHisSerLysAspLysLysLysProGlnArGAspSerGlnAl 108
 Db 1353 -----TATGCCAAGCTC----- 1364
 QY 108 agLInLysSerProValLysSerLysSerThrHisArgIleGlnHisAsnIleAspTyrLe 128
 Db 1365 -----TTCGACTACAT 1375
 QY 128 uLys-----HisLeuSerLysValLysLysIleProSerAspIleGluGlySe 144
 Db 1376 CACACAGAGAAAGTTGCATGTCAGCAACATGGCCAG-----GACAAGAG 1420
 QY 144 rGlyTyrThrAspLeuGlnGluArGlyAspAsnAspIleSerProPheSerGlyAspG1 164
 Db 1421 CGGCTTACAGAGCGTG-----GACTTGGTGAATTCGAC-----AACGAGAA 1462
 QY 164 yGlnProPheLysAspIleProGly-----LysGlyGluAlaThrGlyProAspLeuGluG1 183
 Db 1463 CGAACCAATGCCATATCGCGCTCCCAAGCTGAGGCGGAGGAAAGAGAGAGAGA 1522
 QY 183 yLysAspIleGlnThrGlyPheAlaGlyProSerGluAlaGluSerThrHisLeuAspTh 203
 Db 1523 CGACGAT-----GGCGACTCGAGTGAAGACGCCAGGTGAGACTT 1564
 QY 203 rLysLysProGlyTyrAsnGluIleProGluArGluGluAsnGlyGlyAsnThrIleG1 223
 Db 1565 C---AAGCCCAAGAGAGAGATCCGATGTCGCCGAGGAGATATACAGCAACCGTGAGAG 1621
 QY 223 yThrArGAspGluThrAlaLysGluAlaAspAlaValaLysPylSerLeuValGluGlySe 243
 Db 1622 TGATTGCGACGAT-----GACAGGAGATGCTAGTGC-----GGCGG 1657

```

1 OPERATING SYSTEM: PC-DOS/MS-DOS
2 SOFTWARE: Patent In Review #1.0, Version #1.25
3 CURRENT APPLICATION DATA:
4 APPLICATION NUMBER: PCT/US92/11107
5 FILING DATE: 19921218
6 CLASSIFICATION:
7 PRIORITY APPLICATION DATA:
8 APPLICATION NUMBER: US 07/539,906
9 FILING DATE: 18-JUN-1990
10 ATTORNEY/AGENT INFORMATION:
11 NAME: Granahan, Patricia
12 REGISTRATION NUMBER: 32,227
13 REFERENCE/DOCKET NUMBER: MTT-4787AAA
14 TELECOMMUNICATION INFORMATION:
15 TELEPHONE: 617-861-6240
16 TELEFAX: 617-861-9540
17 INFORMATION FOR SEQ ID NO: 10:
18 SEQUENCE CHARACTERISTICS:
19 LENGTH: 2384 base pairs
20 TYPE: NUCLEIC ACID
21 STRANDEDNESS: double
22 TOPOLOGY: linear
23 MOLECULE TYPE: cDNA
24 ORIGINAL SOURCE:
25 ORGANISM: Drosophila melanogaster
26 IMMEDIATE SOURCE:
27 CLONE: Drosophila SSRP - composite sequence
28 POSITION IN GENOME:
29 CHROMOSOME/SEGMENT: 2
30 MAP POSITION: 60A 1-4
31 FEATURE:
32 NAME/KEY: CDS
33 LOCATION: 123..2291
34 Pct-US92-11107-10
35
36 Alignment Scores:
37 Pred. No.: 0.000345 Length: 2384
38 Score: 138.50 Matches: 99
39 Percent Similarity: 31.91% Conservative: 50
40 Best Local Similarity: 21.20% Mismatches: 136
41 Query Match: 6.08% Indels: 182
42 DB: 5 Gaps: 23
43
44 US-09-700-696B-2 (1-430) x PCT-US92-11107-10 (1-2384)
45
46 QY 16 GlyLeuArgMetSerIleTyrProLysSerThrGly-AsnLysGlyPheGluAspGlyAs 35
47 ||| ||||| : : : : : ||||| : : : : : ||||| : : : : :
48 Db 1187 GGAGCGAGGAGTATCTATATATCCACAAAGCCACCGCTGCATATCCGCTTTGAGAG- 1241
49
50 QY 35 PASPALILeserLysLeuHISAspGlnGluIuTyrGlyALA- 49
51 ||||| : : : : : ||||| : : : : :
52 Db 1242 -----ATTAGTTCGTGTGAACCTTCCGCCGACCGCGGATCCACGGCATCTTCGACTT 1294
53
54 QY 50 -----AlaLeuILaArgAsnAsnMetGlnHisIleMetGlyProValIThrAlaIleLysLe 68
55 ||||| : : : : : ||||| : : : : :
56 Db 1295 CGAAGTACGCTCAAGACGCGAAGCTTTCACATCTTTCCTCCATC- 1340
57
58 QY 68 uLeuGlyGluAsnLysGluAsnThrProArgAsnValLeuAsnIleIleProLase 88
59 ||| : : : |||
60 Db 1341 -----GAGAAGAGAGAG- 1352
61
62 QY 88 mMetAsnTyrAlaLysAlaHisSerLysAspLysLysProGlnArgAspSerGlnAl 108
63 ||||| : : : : :
64 Db 1353 -----TATGCACAGCTC- 1364
65
66 QY 108 agLInLysSerProValLysSerLysSerThrHisArgIleGlnHisAsnIleAspTyrLe 128
67 ||||| : : : : :
68 Db 1365 -----TTCCAGCTACAT 1375
69
70 QY 128 uLys- HisLeuSerLysValLysLysIleProSerAspPheGluGlyLe 144
71 : ||||| : : : ||| : : : ||
72 Db 1376 CACACGACGAAGAGTTCAGTCTGTCAGAACATGGGCGAAG- GACACAGAG 1420
73

```


Db 2055 GAGTC-AGGAGCATCTCTAATCCAAAGATCCTCC-----CTAAAGCT 2098
Oy 95 HisSerLysAspLysLysLysProGlnArg-----AspSerGlnAlaGln 109
Db 2099 CC-AGCAAGAGCCCCCGGGCCCGCCAGACCCCGGAGAGAGAGAGAGAGAGAG 2157
Oy 110 LysSerProValLysSerLysSerThrHisArgIleGlnHisAsnIleAspTyrLeuLys 129
Db 2158 AAGTCTCCGGGACAGCAG 2217
Oy 130 HisLeuSerLysValLysLysIleProSerAspPheGluLysSerGlyTyrThrAspLeu 149
Db 2218 -----AAACCTGTCAGAGCTCTCCGCGGAGAGTTCAGGAGCAGCAGCTG 2262
Oy 150 GlnGluArgLysAspAsnAspIleSerProPheSerGlyLysPglLysIleProPheLysAsp 169
Db 2263 CAGGGGAG 2322
Oy 170 IlePro-----GlyLysGlyGlu----- 175
Db 2323 AAGCCCAAGGTGAAG 2382
Oy 176 AlaThrGlyProAspLeuGluLysAspIleGlnThrGlyPheAlaGlyProSerGlu 195
Db 2383 GCAAGTCCCGCCCTCCAGTGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2442
Oy 196 AlaGluSer-----ThrHisLeuAspThrLysLysProGlyTyr----- 208
Db 2443 GCTCTCTCAGAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2502
Oy 209 -----AsnGluIleProGluArgGluLysAsnGlyLysAsnThr 221
Db 2503 GCTCTGTTCCTCCAGTGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2562
Oy 222 IleGlyThrArgAspLysValLysGluAlaAspAlaValAspValSerLeuValGlu 241
Db 2563 AGTGGCTGCGCCAG 2622
Oy 242 GlySerAspAspIleMetLysSerThrAspPheLysGluLeuProGlyArgGluLysAsn 261
Db 2623 CTTTGAAG 2682
Oy 262 ArgVal-----AspAlaGlySerGlnAsnAla 270
Db 2683 ATGGTGAAGATCAGCCAGTGTCTCTCGGATACCCAGAGAGAGAGAGAGAGAG 2742
Oy 271 HisGlnGlyLysValGluPheHisLysThrProProAlaProSerLysGluLysGlu 290
Db 2743 CGCCAG 2802
Oy 291 GlySerAspAlaAlaGluSerThrAsnThrAsnGluIleProLysAsnGlyLysGly 310
Db 2803 AGGAGCTCAG 2847
Oy 311 SerThrArgLysGlyValAspHisSerAsnArgAsnGlnAlaThrLysAsnGluLysGln 330
Db 2848 GAAGCAG 2880
Oy 331 ArgPheProSerLysGlyLysSerGlnGlyLeuProIleProSerArgGlyLeuAspAsn 350
Db 2881 AGA-----CTGCAAGAG 2892
Oy 351 GluIleLysAsnGluMetAspSerPheAsnGlyProSerHisGluAsnIleLeuThr 369
Db 2893 GAATTCAAATACAGTATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 392
Oy 370 -----HisGlyArgLysLysLysLysLysLysLysLysLysLysLysLys 382
Db 2953 CCCTCAGAGCCCTCTCAGAGTCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3012
Oy 383 AsnSerThrArgAsnLysGlyMetProGlnGlyLysGlySer----- 396

Db 3013 TCGTCTCCAG 3072
Oy 397 TysGlyArgGlnPro-----HisSerAsnArgPheSerSer 409
Db 3073 TGTGGCCAG 3132
Oy 410 ArgArgArgAspAspSerSerSerSerSerSerSerSerSerSerSerSerSerSer 426
Db 3133 ATCCCAACAGCAG 3192
Oy 427 SerAspGlyAsp 430
Db 3193 TCTTCGAGAGAT 3204

RESULT 8
US-08-545-860D-27
; Sequence 27, Application US/08545860D
; Patent No. 6040140
; GENERAL INFORMATION:
; APPLICANT: Croce, Carlo
; APPLICANT: Canaan, Eli
; TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
; TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
; TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the ALL-1 Region
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
; ADDRESSEE: NO. 6040140-15
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/545,860D
; FILING DATE: 07-MAR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/04496
; FILING DATE: 22-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/10930
; FILING DATE: 09-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/327,392
; FILING DATE: 19-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/320,559
; FILING DATE: 11-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/062,443
; FILING DATE: 14-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/971,094
; FILING DATE: 30-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/888,839
; FILING DATE: 27-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/805,093
; FILING DATE: 11-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca Esq., Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: TJD-1262
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439

NAME: Deluca Esq., Mark
 REGISTRATION NUMBER: 33,229
 REFERENCE/DOCKET NUMBER: JTU-1242
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (215) 568-3100
 TELEFAX: (215) 568-3439
 INFORMATION FOR SEQ ID NO: 27:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 9370 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 469..4032
 PCT-US94-04496-27

Alignment Scores:
 Pred. No.: 0.00378
 Score: 137.00
 Percent Similarity: 100
 Best Local Similarity: 34.704
 Query Match: 21.554
 Mismatches: 61
 Indels: 192
 Gaps: 113
 DB: 5

US-09-700-696b-2 (1-430) x PCT-US94-04496-27 (1-9370)

37 AAlaIleSerLysLeuHisAspGlnGluGlyAlaAlaLeuIleArgAsnAsnMet 56
 1938 GCAGCTGACACATGCTGCTACCAAGTCCAGCCAGCTGCGCCACACAGAGGCCAG 1997
 57 GlnHisIleMetGlyProValThrAlaIleLysLeuGluGlyGluAsnLysGln 76
 1998 GAGCAC---AGAGCCCGCCAGCGGCGACCCAGAGTAAGTAAAGCAGCAGCAGTCCAC 2054
 77 ThrProArgAsnValLeuAsn-----IleIleProAlaSerMetAsnTyrAlaLysAla 94
 2055 GAGTC-AGGAGCATCTGATCATCAAGATCCCTCCC-----CTAAAGCT 2098
 95 HisSerLysAspLysLysLysProGlnArg-----AspSerGlnAlaGln 109
 2099 CC-ACCAAAAGCCCGCCCGCCCGCCAGACCCCGCCAGAAAGAGAGCTGTACG 2157
 110 LysSerProValLysSerLysSerThrHisArgIleGlnHisAsnIleAspTyrLeuLys 129
 2158 AAGTCTCCGCGACAGCAGAGAGCCCGCCACAAAGCAACCGTTGAACCAACCAACCA 2217
 130 HisLeuSerLysValLysLysIleProSerAspPheGluGlySerGlyTyrThrAspLeu 149
 2218 -----AAACCTGCAAGGCTCTGCGCGGAGAGTTACGAGCAGCAGCTG 2262
 150 GlnGlnArgGlyAspAsnAspIleSerProPheSerLysAspGlnProPheLysAsp 169
 2263 CAGGGGGAAGAGGAGCCAGGCTTCTTCTCTATGCTCCGAGACCAAGCTTCCAAAGAC 2322
 170 IlePro-----GlyLysGlyGlu----- 175
 2323 AAGCCCAAGTGAAGAGCAAGAGCGCCCGCGCCAGCAAGCAAGCAAGCAAGCAAGCA 2382
 176 AlaThrGlyProAspLeuGluGlyLysAspIleGlnThrGlyPheAlaGlyProSerGln 195
 2383 GCAAGTCCCGCCCTCCAGTGAAGAAAGACACAAAGAGTCCCTCCCTCCCTCCTAG 2442
 196 AlagLysLys-----ThrHisLeuAspThrLysLysProGlyTyr--- 208
 2443 GCTCTCTAGGCGCCAGAACCCGCGAGAGCAATGTGAGAGACAGAGCCCTAGCACTTT 2502
 209 -----AsnGluIleProGlnArgGluGluAsnGlyLysAsnThr 221
 2503 GCTCTGTTCCCTGACTAGAGAGGCGCCAGCCACAGTGTGAGGCGGAGGAGT 2562
 222 IleGlyThrArgAspGluThrAlaLysGluAlaAspAlaValAspValSerLeuValGlu 241

2563 AGTGGCTCCCGCCCAAGCCGTGTGCTCCAGAGAGACCGCGCAAGACAGACTCCCATTTG 2622
 242 GlySerAsnAspIleMetGlySerThrAsnPheLysGluLeuProGlyArgGluLysAsn 261
 2623 CCTTGAAGAGACACCAAGTGTCTCAGCGCTCAGAGGACACTCTCTCCCAACCAAGCTTG 2682
 262 ArgVal-----AspAlaGlySerGlnAsnAla 270
 2683 ATGTAACATACACCTAGACCTGCTCTCGATATACCCAGCTCCGGAGAGGAGGC 2742
 271 HisGlnGlyLysValGluPheHisTyrProPheAlaProSerLysGluArgLysGlu 290
 2743 CGCCAGAGAGAAAGAGATTAACAGCCCGCCGAGGAAAGACACAGCTGTGAGAG 2802
 291 GlySerSerAspAlaIleGluSerThrAsnTyrAsnGluIleProLysAsnGlyLysGly 310
 2803 AGGACTCAGACAGCTCA-----AGCAAGTTGGCCAAAGCAAGAAAGGT 2847
 311 SerThrArgLysGlyValAspHisSerAsnArgAsnGlnAlaThrLeuAsnGluLysGln 330
 2848 GAAGCAGAAAGAGACTGTAT-----AACAGAAATC 2880
 331 ArgPheProSerLysGlyLysSerGlnGlyLeuProIleProSerArgGlyLeuAspAsn 350
 2881 AGA-----CTGGAGAG 2892
 351 GluIleLysAsnGluMetAspSerPheAsnGlyProSerHisGluAsnIleIleThr--- 369
 2893 GAATTCAAATACAGTATCTTCTTCATCTCCACCAAAACAACTCTAAACCAAG 2952
 370 -----HisGlyArgLysTyrHisTyrValProHisArgGlnAsn 382
 2953 CCTCCAGGCCCTCTCAGCTCAGCTCTCAAGAGAAATGCTCCCGCCAGCCGTGCTC 3012
 383 AsnSerThrArgAsnLysGlyMetProGlnGlyLysGlySer----- 396
 3013 TCGTCCCTCCAGAGACCGCAAGCCTGCACTTAAGAGTCAAGCGCGAGACAGACACC 3072
 397 TrpGlyArgGlnPro-----HisSerAsnArgArgPheSerSer 409
 3073 TGTGGCCAGAGCCCTCCCAAAAGTGCAGCAGATACCAAGACCAACCAACCAACTTTC 3132
 410 ArgArgArgAspSerSerSerGluSerSerAspSerGlySerSerSerGlu----- 426
 3133 ATTCACAGAGAGAGAGTGAAGGAGGAGGCTCCAGAACTCTCGAGACCAAGGT 3192
 427 SerAspGlyAsp 430
 3193 TCTTCGAGAT 3204

RESULT 10
 US-08-320-559-25
 Sequence 25, Application US/08320559
 Patent No. 5633135
 GENERAL INFORMATION:
 APPLICANT: Croce, Carlo
 APPLICANT: Canaan, Eli
 TITLE OF INVENTION: Diagnostics, Therapeutics and Methods for
 TITLE OF INVENTION: Detection and Treatment of Acute Leukemias
 TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the
 NUMBER OF SEQUENCES: 44
 CORRESPONDENCE ADDRESS:
 ADDRESSSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5633135r1s
 STREET: One Liberty Place - 46th Floor
 CITY: Philadelphia
 STATE: PA
 COUNTRY: USA
 ZIP: 19103
 COMPUTER READABLE FORM:
 MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
 COMPUTER: IBM PS/2

Patent No. 6040140
 GENERAL INFORMATION:
 APPLICANT: Croce, Carlo
 APPLICANT: Canaan, Eli
 TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
 TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
 NUMBER OF SEQUENCES: 94
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Woodcock, Washburn, Kurtz, Macklewitz &
 ADDRESSEE: No. 6040140r1s
 STREET: One Liberty Place, 46th floor
 CITY: Philadelphia
 STATE: Pennsylvania
 COUNTRY: USA
 ZIP: 19103
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/545,860D
 FILING DATE: 07-MAR-1996
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/US94/04496
 FILING DATE: 22-APR-1994
 APPLICATION DATA:
 APPLICATION NUMBER: PCT/US92/10930
 FILING DATE: 09-DEC-1992
 APPLICATION DATA:
 APPLICATION NUMBER: US 08/327,392
 FILING DATE: 19-OCT-1994
 APPLICATION DATA:
 APPLICATION NUMBER: US 08/320,559
 FILING DATE: 11-OCT-1994
 APPLICATION DATA:
 APPLICATION NUMBER: US 08/062,443
 FILING DATE: 14-MAY-1993
 APPLICATION DATA:
 APPLICATION NUMBER: US 07/971,094
 FILING DATE: 30-OCT-1992
 APPLICATION DATA:
 APPLICATION NUMBER: US 07/888,839
 FILING DATE: 27-MAY-1992
 APPLICATION DATA:
 APPLICATION NUMBER: US 07/805,093
 FILING DATE: 11-DEC-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: Deluca Esq., Mark
 REGISTRATION NUMBER: 33,229
 REFERENCE/DOCKET NUMBER: TJU-1262
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (215) 568-3100
 TELEFAX: (215) 568-3439
 INFORMATION FOR SEQ ID NO: 25:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 9391 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 421..4053
 US-08-545-860D-25

Alignment Scores:
 Pred. No.: 0.00379 Length: 9391
 Score: 137.00 Matches: 100
 Percent Similarity: 34.708 Conservative: 61
 Best Local Similarity: 21.558 Mismatches: 192

Query Match: 6.01% Indels: 113
 DB: 3 Gaps: 17
 US-09-700-696b-2 (1-430) x US-08-545-860D-25 (1-9391)
 QY 37 AAlleSerLysLeuHisAspGlnGluGlyrGlyAlaLeuLeuLeuArgAsnMet 56
 DB 1959 GCAGCTGCACAACTGGCTGACCAAGTCCAGCCAGCCAGCTGCCACCAAGAGGCCCGCAG 2018
 QY 57 GlnHisIleMetGlyProValThrAlaIleLysLeuGluGlyGluValAsnGlyGln 76
 DB 2019 GAGCAC--AGAGCCCCCAGCGCCGACCACAGAGAGTAAGGCGACAGCAGAGTCCAC 2075
 QY 77 ThrProArgAsnValLeuAsn-----IleLeuProAlaSerMetAsnTyrAlaAla 94
 DB 2076 GAGTC-AGAGAGCATTCGATATTCCAAGATCTCCCTCC-----CTAAAGCT 2119
 QY 95 HisSerLysAspLysLysLysLysProGlnArg-----AspSerGlnAlaGln 109
 DB 2120 CC-ACCAAGAGCCCCCGGGCCGCCACCCGAGAGCCCCCAGAGAGAGAGAGCTGTCAG 2178
 QY 110 LysSerProValLysSerLysSerThrHisArgIleGlnHisAsnIleAspTyrLeuLys 129
 DB 2179 AAGTCTCCGGCAGCAGCAGAGAGCCCCCAGAGGCAAAACGTTGGACCAAAACCAACCCAA 2238
 QY 130 HisLeuSerLysValLysLysIleProSerAspPheGluGlySerGlyTyrThrAspLeu 149
 DB 2239 -----AAACGTGCAAGCCCTGCGCGGCGACAGTTACGAGGACAGCTG 2283
 QY 150 GlnGluArgLysAspAsnAspIleSerProPheSerGlyAspGlnProPheLysAsp 169
 DB 2284 CAGGGGGAAGAGGAGCAGAGGCTTCTTCTGATGCTCCGAGACAGACTTCCAAAGAC 2243
 QY 170 IlePro-----GlyLysGlyGlu----- 175
 DB 2244 AAGCCCAAGGTGAAGAGCAAGAGAGCGCCCGCCGAGCAAGCAAGCAAGCAAGCAAGCA 2403
 QY 176 AlaThrGlyProAspGlnGluGlyLysAspIleGlnThrGlyPheAlaGlyProSerGlu 195
 DB 2404 GCAGTGCCTCCCTCCAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2463
 QY 196 AlaGluSer-----ThrHisLeuAspThrLysLysProGlyLys 208
 DB 2464 GCTCTCTCAGGCCAGAGACCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2523
 QY 209 -----AsnGluIleProGluArgGluGluGluAsnGlyLysAsnThr 221
 DB 2524 GCTCTTGTTCCTCCCTGACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2583
 QY 222 IleGlyThrArgAspGluThrAlaLysGluAlaAspAlaValAspValSerLeuValGlu 241
 DB 2584 AGTGGTCCGCGCAGAGCGGTGTGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2643
 QY 242 GlySerAsnAspIleMetGlySerThrAsnPheLysGluLeuProGlyArgGluGlyAsn 261
 DB 2644 CTTTGAGAGAGACCAAGCTGCTCCACCGCTCAGAGAGAGAGAGAGAGAGAGAGAGAG 2703
 QY 262 ArgVal-----AspAlaGlySerGlnAla 270
 DB 2704 ATGTGAGAGATCACCCCTGAGACTGCTCTCTCGGATACCCCAAGCTCCCGGAGAGAGAG 2763
 QY 271 HisGlnGlyLysValGluPheHisTyrProProAlaProSerLysGluLysArgLysGlu 290
 DB 2764 CGCCAGAGAGAGAGAGAGATTAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2823
 QY 291 GlySerSerAspAlaAlaGluSerThrAsnTyrAsnGluIleProLysAsnGlyLysGly 310
 DB 2824 AGGAGCTCAGACAGCTCA-----AGCAAGTTGGCCCAAAAGAGAGAGAGAGAGAG 2868
 QY 311 SerThrArgLysGlyValAspPheHisSerAsnArgAsnGlnAlaThrLeuAsnGluLysGln 330
 DB 2869 GAACGAGAGAGAGAGAGAGCTGTGAT-----AACAGAGAAATTC 2901


```

Db 2824 AGGAGCTCAGACAGCTC-----AGCAAGTTGGCCAAAAGAAAGGCT 2868
QY 311 SerThrArgLysGlyValAspHisSerAsnArgAsnGlnAlaThrLeuAsnGlnLysGln 330
Db 2869 GAACGAGAAAGAGACTGTGAT-----AACCAAGAAATC 2901
QY 331 ArgPheProSerLysGlyLysSerGlnGlyLeuProIleProSerArgGlyLeuAspAsn 350
Db 2902 AGA-----CTGAGAGAG 2913
QY 351 GluIleLysAsnGluMetAspSerPheAsnGlyProSerHisGluAsnIleIleThr--- 369
Db 2914 GAATCAATATCAGATCATCTTCATCTTCATCCCAACAAGATCTCTTAAACAAG 2973
QY 370 -----HisGlyArgLysThrLysThrValProHisArgGlnAsn 382
Db 2974 CCTTCAGGCGCTCTCAGCTCTCAAGAGAAATGCTCCCGCCCGCCGCTGCTC 3033
QY 383 AsnSerThrArgAsnLysGlyMetProGlnGlyLysGlySer----- 396
Db 3034 TCGTCTCCCGAGAGCCAGCCAGCCAGCTGACTTAAGAGATCAAGCGGAGAGACACC 3093
QY 397 TrpLysArgGlnPro-----HisSerAsnArgArgPheSerSer 409
Db 3094 TGTGCGCAGGACCTCCCAAAAGTCCAGACAGTACCAAGAGACCAACAAGACTCTTCC 3153
QY 410 ArgArgArgAspAspSerSerGluSerSerAspSerGlySerSerGlu----- 426
Db 3154 ATTCGCCAGCAGAGAGATAGAGGAGGAGGCTCCAGAAAGCTCTCTGAGACCAAGGCT 3213
QY 427 SerAspLysAsp 430
Db 3214 TCTTCCGGAGAT 3225

RESULT 13
US-08-568-459A-1
; Sequence 1, Application US/08568459A
; Patent No. 5849306
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnals, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08-568,459A
; FILING DATE: 07-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelson, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 4084 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium vivax
; US-08-568-459A-1

Alignment Scores:
Pred. No.: 0.00194 Length: 4084
Score: 134.50 Matches: 97
Percent Similarity: 36.428 Conservative: 80
Best Local Similarity: 19.968 Mismatches: 220
Query Match: 5.90% Indels: 89
DB: 2 Gaps: 21

US-09-700-696B-2 (1-430) x US-08-568-459A-1 (1-4084)

QY 2 AsnLysGlnLysSerLysSerAsnLysGluAsnThrHisAsnGlyLeuArgMetSerIle 21
Db 1724 AATTAATTCATATAAGTGTAAAAAGCGAAGAGGTTGACAGCGGACAGTATCTGTAACCTCT 1783
QY 22 TyrProLysSerThrGlyAsnLysGlyPheGlnAspGlyAspAspAlaIleSerLysLeu 41
Db 1784 TAT-----GATTAATAAAACG--- 1801
QY 42 HisAspGlnGlnLysGlyAlaAlaLeuIleArgAsnAsnMetGlnHisIleMetGly 61
Db 1802 ---GAGTTAGATGATTAATTAACGAGTGGCTTTGAGAGATGAATTAACAAGTGATGCT 1858
QY 62 ProValThrAlaIleLysLeuLeuGly---GluGluAsnLysGluAsnThrProArgAsn 80
Db 1859 GCATATATTGAGTTATGCTGTTGCTTCCGTTGAAGAGCTTAATAAAATCTCAGGAAGTT 1918
QY 81 ValLeuAsnIleIleProAlaSerMetAsnTyrAlaLysAlaHisSerLysAspLysLys 100
Db 1919 GTGCAAAATGTGCAAAATGCTGTAAATCTCAGGCC---ACCAATTCAAATCCGATAGT 1975
QY 101 LysProGlnArgAspSerGlnAlaGlnLysSerProValLysSerLysSerThrHisArg 120
Db 1976 CAGCCTGTAGATAGTAAAGCGGAGAGGTTCCA-----CGAGATTCATACGAT--- 2026
QY 121 IleGlnHisAsnIleAspTyrLysHisLysLeuSerLysVal---LysLysIleProSer 139
Db 2027 ---GAAATGTTAAACAGTGGCCAGATAGTCTACACAGGTAAACCTTTACGGGG 2080
QY 140 AspPheGlnGlySerGlyTyrThrAspLeuGln-----GluArgGlyAspAsnAsp 156
Db 2081 GATGTCGCAAAATGGAATCAGACACCTGCAAAAGCATGTACACGAAAGTATATTGCC 2140
QY 157 IleSerProPheSerGlyAspGlyGlnProPheLysAspIleProGlyLysGlyGlu--- 175
Db 2141 GAAAGTGAAGTGTAAATGTTGATCCGAGAAATCTGTAAATGAAGATGATACACAC 2200
QY 176 -----AlaThrGlyProAspLeuGlnGlyLysAspIleGlnThrGlyPheAlaGly 192
Db 2201 ACTGCAAGGCTTACAGATATTCCGAGCTGGAAGGAA---AAGTTAGGCCATCAAT 2257
QY 193 ProSerGluAlaGluSerThrHisLeuAspThrLysLysProGly----- 207
Db 2258 AGTCGACCTTCGAGTCCACC---GTTGAAGCAATATACCCAGGTGATGATCTGTAAC 2314
QY 208 -----TyrAsnGluIle 211
Db 2315 AGTCATCATATCTGTAGTGAAGTGTGTAACCAATTCGTAACCCCTTAATAAGTTTG 2374
QY 212 ProGluArgGlnGluAsnGlyLysAsn-----Thr 221
Db 2375 AGCATTCGAAAGACAAATAGATGATGATGACCTGGGGAATCAATGCGAATCTGAT 2434
QY 222 IleGlyThrArgAspGluThrAlaLysGlnAlaAspAlaValAspValSerLeuValGlu 241

```

```

Db 2435 TCAAAATAGTAAGGTGAGACGGGAAAGCGCAAGATATATATGCGGAGGCTACTAA 2494
      ::::: ||||| |||
Qy 242 GlycerasnaSpilleMetGlySerThrAsnPhelGluLeuProGlyArgGlyAsn 261
      ||::: ||:::
Db 2495 GATAGTGTATATGTTGAGATGGTACAGC-----TCTGCTACGGGTGAT 2539
      ||::: ||:::
Qy 262 ArgValAspAlaGlySerGlnAsnAlaHisGlnGlyValGluPheHisTyrProPro 281
      ||||| ::::: |||
Db 2540 ACTACTGATCAGTTGATGAGAAATTAATTAAGGCTTCTCTGAGGATGAGATAAACT 2599
      ||||| ::::: |||
Qy 282 AlaProSerGlyGluValArgGlyGlySer-----AspAlaAlaGluSer 298
      ||||| ||||| |||
Db 2600 GTAGGAGTAAAGATGAGGGGGGGAAGATACCTCTCAATAAGATGACGACACTGA 2659
      ||||| ||||| |||
Qy 299 ThrAsnTyrAsnGluLeuProLysAsnGlyLysGlySerThr-----ArgLysGly 315
      ::::: ||| |||
Db 2660 GTTGGTGAAGTATGATTCGTGAGACACAGCGCTGTGTGACACTAATGATGATCAAA 2719
      ::::: ||| |||
Qy 316 ValAspHisSerAsnArgAsnGlnAlaThrLeuAsnGluLysGlnArg-----PhePro 333
      ||| ::::: |||
Db 2720 AATGACACGGAAAGAAAGCGGCGCTGACCTGACAGTAAACAAAGTGAAGATCAACT 2779
      ||| ::::: |||
Qy 334 SerLysGlyLysSerGlnGlyLeuProIleProSerArgGly-----LeuAspAsnGlu 351
      ::::: ||||| |||
Db 2780 GCGCTAAGTAAACCGAAAGTTAGATCAACAGAAAGTGAGATGACACTAATGAT 2839
      ::::: ||||| |||
Qy 352 IleLysAsnGluMetLysSerPheAsnGlyProSerHisGlnAsnIleIleThrHisGly 371
      ||| ::::: |||
Db 2840 ACACTAAGTGTAGAAAATAAATAAAGGAAAGAAAGAAAGGATTTACAAACCATGAT 2899
      ||| ::::: |||
Qy 372 ArgLysTyrHisTyrValProHisArgGlnAsnSerThrArg----- 386
      ||| ::::: |||
Db 2900 TTTAAAGTAATGATACGCCGATGAAGAAACCAATTTCTGTCAATACAGATGACGAA 2959
      ||||| ::::: |||
Qy 387 -----AsnLysGlyMetProGlnGlyLysGlySerTyrGlyArgGlnProHisSer 403
      ||||| ::::: |||
Db 2960 GGACATGACAGGATAGCATCAAAATGATTAAGCAGAA-----AGAGAAACCATATG 3013
      ||||| ::::: |||
Qy 404 AsnArg---ArgPheSerSerArgArgArgAspSerSerSerGlySerSerAspSerGly 422
      ||||| ||||| |||
Db 3014 AATTAAGTACTTTTACGAAAATAACAAATAGTCACCATTTAAATGATTAATTAATTG 3073
      ||||| ||||| |||
Qy 423 SerSerSerGlySerAsp 428
      ||||| ||||| |||
Db 3074 AGTAATGAAAATTAATGAT 3091
      ||||| ||||| |||

RESULT 14
US-08-487-826B-1
: Sequence 1, Application US/08487826B
: Patent No. 5993827
: GENERAL INFORMATION:
: APPLICANT: Sim, Kim L.
: APPLICANT: Chitnis, Chetan
: APPLICANT: Miller, Louis H.
: APPLICANT: Peterson, David S.
: APPLICANT: Su, Xin-zhaun
: APPLICANT: Wellem, Thomas E.
: TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
: TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
: NUMBER OF SEQUENCES: 45
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Knobb Martens Olson & Bear
: STREET: 620 Newport Center Drive 16th Floor
: CITY: Newport Beach
: STATE: California
: COUNTRY: US
: ZIP: 92660
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25

```

```

: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/487,826B
: FILING DATE: 10-SEP-1993
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Israelson, Ned
: REGISTRATION NUMBER: 29,655
: REFERENCE/DOCKET NUMBER: NIH121,001CPI1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 235-8550
: TELEFAX: (619) 235-0176
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4084 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEICAL: NO
: ORIGINAL SOURCE:
: ORGANISM: Plasmodium vivax
: US-08-487-826B-1

Alignment Scores:
Pred. No.: 0.00194 Length: 4084
Score: 134.50 Matches: 97
Percent Similarity: 36.42% Conservative: 80
Best Local Similarity: 19.96% Mismatches: 220
Query Match: 5.90% Gaps: 89
Db: 2 Gaps: 21

US-09-700-696b-2 (1-430) x US-08-487-826B-1 (1-4084)
Qy 2 AsnLysGlyTyrSerIleSerAsnLysGluAsnThrHisAsnGlyLeuArgMetSerIle 21
      ||||| ||||| |||
Db 1724 AATTAATTCATTAAGTATAAATAAAGCGAAAGGTTGACAGCGGAGGTATGTAACCTG 1783
      ||||| ||||| |||
Qy 22 TyrProLysSerThrGlyAsnLysGlyPheGluAspGlyAspAlaIleSerLysLeu 41
      ||| ||||| |||||
Db 1784 TAT-----GATATATCTAAACAG- 1801
      |||
Qy 42 HisAspGlnGluGlyTyrGlyAlaIleLeuIleArgAsnAsnMetGlnHisIleMetGly 61
      ::::: ||||| |||
Db 1802 ---GAGTGAATGAAATTTAACGAGGTGGCTTTGAGATGAATTAACAACGATGATGCT 1858
      ::::: ||||| |||
Qy 62 ProValThrAlaIleLysLeuGly---GluGluAsnLysGluAsnThrProArgAsn 80
      ||| ::::: ||||| |||
Db 1859 GCATATATGATGATGCGTTGTCCTGTAAGAGGCTAAATAAATACTCAGGAAGTT 1918
      ||| ::::: ||||| |||
Qy 81 ValLeuAsnIleIleProAlaSerMetAsnTyrTrlAlaLysAlaHisSerLysAspLysLys 100
      ||| ||||| ||||| |||
Db 1919 GTGACAAATGTGACAAATGCTGTAATCTCAGGC---ACCAATTCAATCCATAGT 1975
      ||| ||||| ||||| |||
Qy 101 LysProGlnArgAspSerGlnAlaGlnLysSerProValLysSerLysSerThrHisArg 120
      ::::: ||||| ||||| |||
Db 1976 CACCTGTAGTACTAGTAAGACGAGAGAGTTCCA-----GGAGATTCTACGCAT--- 2026
      ::::: ||||| ||||| |||
Qy 121 IleGlnHisAsnIleAspTyrLeuLysHisLeuSerLysVal---LysLysIleProSer 139
      ||||| ||||| ||||| |||
Db 2027 ---GGAATGTTAACTGACAGTGGCCAAAGATAGTTCTACACAGGTAAAGCTGTACGGGG 2080
      ||||| ||||| ||||| |||
Qy 140 AspPheGluGlySerGlyTyrThrAspLeuGln-----GluArgGlyAspAsnAsp 156
      ||| ::::: ||||| |||
Db 2081 GATGCTCAAAATGGAATCAACACCTGCAAGAACCATGACGACGAGATATGCTC 2140
      ||| ::::: ||||| |||
Qy 157 IleSerProPheSerGlyAspGlyGlnProPheLysAspIleProGlyLysGlyGlu--- 175
      ||| ::::: ||||| |||
Db 2141 GAAAGTGAAGTGCATAAATAATGTTGATCCGACAAATCTGTAAGTAAAGAGAGTACAGAC 2200
      ||| ::::: ||||| |||
Qy 176 -----AlaThrGlyProAspLeuGluGlyLysAspIleGlnThrGlyPheAlaGly 192
      ||||| ||||| ||||| |||
Db 2201 ACTGCACGCTTACAGGATATGCGCAAGCTGGAAGAA---AACTTAGCGCATCAAT 2257
      ||||| ||||| ||||| |||

```

```

QY 193 ProSerGluAlaGluSerThrHisLeuAspThrLysLysProGly----- 207
DB 2258 AGTCAGACCTTCTGAGTCCACC---GTGAAAGCAAAATAGCCAGGTGATGACTGTCAAC 2314
QY 208 -----TyrAsnGluIle 211
DB 2215 AGTCATCTTATACCTGAGTGTGAAACCATTGTGTAACCCCTTATATGTTTG 2374
QY 212 ProGluArgGluGluAsnGlyLysAsn-----Thr 221
DB 2375 AGGCATTGCAAAAGCAATATGATGATGACGATGACCTGCGGAATCAATGGCGAATCTGTAT 2434
QY 222 ILGGLYThrArgAspGluThrAlaLysGluAlaAspAlaValAspValSerLeuValGlu 241
DB 2435 TCATAATAGTAAAGGTGAGACGGGAGGCGCAAGATAATGATATGGCGAAGCTCTAA 2494
QY 242 GlySerAsnAspIleMetGlySerThrAsnPheLysGluLeuProGlyArgGluLysAsn 261
DB 2495 GATAGTAGTAATAGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2539
QY 262 ArgValAspAlaGlySerGluAsnAlaHisGlnGlyLysValGluPheHisLysProPro 281
DB 2540 ACTACTGATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2599
QY 282 AlaProSerLysGluLysArgLysGluLysSer-----AspAlaAlaGluSer 298
DB 2600 GTAGAGAGTAAAGATGAGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2659
QY 299 ThrAsnLysArgGluLysProLysAsnGlyLysGlySerThr-----ArgLysGly 315
DB 2660 GTTGGTGAAGTAAATCTGAGAACAGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2719
QY 316 ValAspHisSerAsnArgAsnGlnAlaLeuHisLeuAsnGlnLysGlnAlaArg-----PhePro 333
DB 2720 AATGACACGAGAAAGAACGGGCGCTTACCTGACCTGACGATAAACAAAGTGAAGATGCAACT 2779
QY 334 SerLysGlyLysSerGlnGlyLeuProLysProSerArgLys-----LeuAspAsnGlu 351
DB 2780 GCGCTAAGTAAACCGAAAGTTTGTAGATCAACAGAAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2839
QY 352 IleLysAsnGluMetAspSerPheAsnGlyProSerHisGluAsnIleLeuHisGly 371
DB 2840 ACACTAAACAGTTTAAAGAAATTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2899
QY 372 ArgLysLysHisLysValProHisArgGlnAsnSerSerThrArg----- 386
DB 2900 TTTTAAAGTAAATGATACCGCGATGAAGAACCAATTCATGATCAAACTACAGTGCAGAA 2959
QY 387 -----AsnLysGlyMetProGlnGlyLysGlySerTrpGlyArgGlnProHisSer 403
DB 2960 GGACATGACAGGATGACATCAAAATGATTAAGCAGAA-----AGGAGAAAGCATATG 3013
QY 404 AsnArg---ArgPheSerSerArgArgAspAspSerSerGluSerSerSerSerGly 422
DB 3014 AATTAAGATACCTTTTACGAAATAATACAAATGATGATGATGATGATGATGATGATGATGAT 3073
QY 423 SerSerSerGluSerSer 428
DB 3074 AGTAATGAGAAATTTAGAT 3091

```

RESULT 15

US-09-210-288-1
Sequence 1, Application US/09210288
Patent No. 6392026

GENERAL INFORMATION:

APPLICANT: Shim, Kim L.

APPLICANT: Chilton, Chetan

APPLICANT: Miller, Louis H.

APPLICANT: Peterson, David S.

APPLICANT: Su, Xin-zhuo

APPLICANT: Wellens, Thomas E.

TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX

AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS

```

NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESS: Knobb Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/210,288
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH121.1FMDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4084 base pairs
TYPE: nucleic acid
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORGANISM: Plasmodium vivax
US-09-210-288-1
Alignment Scores:
Pred. No.: 0.00194 Length: 4084
Score: 134.50 Matches: 97
Percent Similarity: 36.42% Conservative: 80
Best Local Similarity: 19.96% Mismatches: 220
Query Match: 5.90% Indels: 89
DB: 4 Gaps: 21
US-09-700-696B-2 (1-430) x US-09-210-288-1 (1-4084)
QY 2 AsnLysGlyLysSerIleSerAsnLysGluAsnThrHisAsnGlyLeuArgMetSerIle 21
DB 1724 AATAAATTCATAGTGTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1783
QY 22 TyrProLysSerThrGlyAsnLysGlyPheGluAspGlyAspAlaIleSerLysLeu 41
DB 1784 TAT-----GATATACATAAAACAG--- 1801
QY 42 HisAspGlnGluGluLysValAlaAlaLeuIleArgAsnAsnMetGlnHisIleMetGly 61
DB 1802 ---GAGTGAATGATTAAGAGAGGCTTTGAGAAATTAATTAACAAACGATGATGCT 1858
QY 62 ProValThrAlaIleLysLeuGly---GluGluAsnLysGluAsnThrProArgAsn 80
DB 1859 GCATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1918
QY 81 ValLeuAsnIleLeuProAlaSerMetAsnTyrAlaLysAlaHisSerLysAspLys 100
DB 1919 GTGACAAATGTCGCAATGCTGCTAATTCACAGCC---ACCAATTCAAATCCGATTAAGT 1975
QY 101 LysProGluArgAspSerGlnAlaGlnLysSerProValLysSerLysSerThrHisArg 120
DB 1976 CACCTGTAGATAGTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2026
QY 121 IleGlnHisAsnIleAspLysLysLysHisLysSerLysVal---LysLysIleProSer 139
DB 2027 -----GAAAGTGTAAACAGTGGCCAGATAGTCTTACACACAGGTAAAGCTGTGTACGGGG 2080

```


GenCore version 5.1.4 p5 4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: April 17, 2003, 00:25:15 ; Search time 122 Seconds

(without alignments)
3540.669 Million cell updates/sec

Title: US-09-700-696b-2

Perfect score: 2279
Sequence: 1 VNKEYSISNKENHNGLRMS.....RRDSSSSSDSSSSSDGD 430

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Zgapop 6.0, Zgapext 7.0
Delop 6.0, Delext 7.0

Searched: 639749 segs, 502280978 residues

Total number of hits satisfying chosen parameters: 1279498

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=ylh
-O=/cgn2.1/USPTO.spool/US09700696/runat_11042003.101727.12478/app.query.fasta.1.583
-DB=Published.Applications_MA -QWt=fastlap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPT=0 -LOOPT=0 -UNITS=bits -START=1 -END=1 -MATRIX=Plowm62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USBR=US09700696 -GCN 1.1.93 -runat_11042003.101727.12478
-NCPU=6 -ICPU=3 -NO_XLPXY -NO_NMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FEAPOP=6
-FEAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published.Applications_MA:*
1: /cgn2.6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2.6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2.6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2.6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
5: /cgn2.6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
6: /cgn2.6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
7: /cgn2.6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2.6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2.6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2.6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2.6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2.6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2.6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2.6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	148	6.5	5721	10	US-09-785-770A-15
2	148	6.5	8121	9	US-09-785-770A-14
3	134.5	5.9	4084	9	US-10-153-273-1
4	133.5	5.9	1960	10	US-09-864-761-4620

5	131	5.7	2025	10	US-09-834-975-822	Sequence 822, App
6	130.5	5.7	1450	9	US-10-001-835-66	Sequence 66, App1
7	129.5	5.7	2454	10	US-09-962-436-563	Sequence 563, App
8	127.5	5.6	4840	10	US-09-880-107-3423	Sequence 3423, App
9	127	5.6	1066	9	US-09-746-660A-63	Sequence 63, App1
10	127	5.6	2154	9	US-09-738-626-2161	Sequence 2161, App
11	127	5.6	5331	10	US-09-764-116-6	Sequence 6, App1
12	125.5	5.5	1975	10	US-09-864-761-3788	Sequence 3788, App
13	125.5	5.5	5173	10	US-09-811-045A-2	Sequence 2, App1
14	125	5.5	2567	9	US-10-149-110-6	Sequence 6, App1
15	123.5	5.4	1374	9	US-10-001-835-71	Sequence 71, App1
16	122.5	5.4	3347	9	US-09-736-457-318	Sequence 318, App
17	122.5	5.4	3347	9	US-09-902-941-318	Sequence 318, App
18	122.5	5.4	3347	9	US-09-849-626-318	Sequence 318, App
19	122.5	5.4	3347	9	US-10-017-754-318	Sequence 318, App
20	122	5.4	5857	9	US-09-932-257A-2	Sequence 2, App1
21	121.5	5.3	1680	9	US-09-298-523B-73	Sequence 73, App1
22	121	5.3	4282	12	US-10-044-090-207	Sequence 207, App
23	120.5	5.3	2110	12	US-10-001-843-60	Sequence 60, App1
24	120.5	5.3	3413	12	US-10-001-843-61	Sequence 61, App1
25	120.5	5.3	4507	9	US-10-153-273-3	Sequence 3, App1
26	120	5.3	1533	9	US-09-894-844-7	Sequence 7, App1
27	120	5.3	3753	10	US-09-801-368-363	Sequence 363, App
28	120	5.3	10383	10	US-09-960-253-181	Sequence 181, App
29	120	5.3	42432	9	US-10-027-806-2	Sequence 2, App1
30	120	5.3	42432	9	US-10-034-623-2	Sequence 2, App1
31	120	5.3	42432	9	US-10-027-801-2	Sequence 2, App1
32	119.5	5.2	3024	9	US-10-211-133-1	Sequence 1, App1
33	119.5	5.2	3024	9	US-10-211-133-6	Sequence 6, App1
34	119.5	5.2	3202	9	US-10-211-133-2	Sequence 2, App1
35	119.5	5.2	43950	10	US-09-834-975-962	Sequence 962, App
36	119.5	5.2	3505	12	US-10-060-332-3	Sequence 3, App1
37	119	5.2	3505	9	US-09-954-531-1360	Sequence 1360, App
38	119	5.2	3505	10	US-09-962-436-272	Sequence 272, App
39	119	5.2	3750	9	US-09-938-842A-1662	Sequence 1662, App
40	118	5.2	6160	9	US-09-902-432-3	Sequence 3, App1
41	117.5	5.2	1714	10	US-09-765-272-3	Sequence 3, App1
42	117.5	5.2	1983	9	US-09-872-462-3	Sequence 3, App1
43	117.5	5.2	2076	9	US-09-986-480-116	Sequence 116, App
44	117.5	5.2	2480	9	US-09-872-462-2	Sequence 2, App1
45	117.5	5.2	3420	9	US-09-872-462-1	Sequence 1, App1

ALIGNMENTS

RESULT 1
US-09-785-770A-15
Sequence 15, Application US/09785770A
Patent No. US20020103360A1
GENERAL INFORMATION:
APPLICANT: Pan, Yang
TITLE OF INVENTION: A NOVEL PROTEIN RELATED TO MELANOMA-INHIBITING PROTEIN
FILE REFERENCE: 07334-328001
CURRENT APPLICATION NUMBER: US/09/785,770A
CURRENT FILING DATE: 2001-02-16
PRIOR APPLICATION NUMBER: US 09/387,462
PRIOR FILING DATE: 1999-09-01
PRIOR APPLICATION NUMBER: US 09/145,056
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn Ver. 4.0
SEQ ID NO 15
LENGTH: 5721
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(5721)
US-09-785-770A-15
Alignment Scores:

[illegible]

Oy	269	AsnIlaHisGInglYlysValGlupheHISTYrPrroProlAlepSerLysGLULYSArg	288
Db	2287	AGACCACTTTTAGGCACCATTCAT-----CCGATCCAGAATTGGAAAGC	2334
Oy	289	LysGIUlyserSeraspIlaInaglUserThrAsnTYrAsnGluIlleProLYsAsngly	308
Db	2335	AAGCAAACAACtAGTATGATTGGATTGTGTAATAAACAAAGTGAG-----	2379
Oy	309	LysGIlyserThraArgLYsgLYvalAspHisSerAsnArgAsnGlnAlaThrLeuAsnGlu	328
Db	2380	---ACTGTCGCCAAAGGGGTCAACACAGGAGCGAGGAACCAAAATRCATGTGTGAA	2433
Oy	329	LysGlnAlaArg-----PheProSerLySolyLysSer	338
Db	2434	AAAAGAACCCCTCTGGCAGATTAAGAAAGCACAGAGACCATTGGAACGAAGTCACCTTTTCT	2493
Oy	339	GInglyLeuProIIeProSerArGgLYleuAspAsnGluIlleLYsAsnGluMetaspSer	358
Db	2494	GACACCATTAATAATTCAAGACTCCGAATTTGGTGAAGTGTTCAGATTAATAATTCGTGAT	2553
Oy	359	Phe-----AsnGlyProSerHisGluAsnIlleIleThrHisGlyArGLysTyrlHisTYr	376
Db	2554	TATCTGAAGAACGACACACCCCTGAGAAACATCTGAAGACTCAAGCGCTT-----	2601
Oy	377	ValProHISArGLINAsnAsnSerThraArgAsnLYseLYmetProGInglyLYsgLYser	396
Db	2602	-----GCAGGGAGGCTTGAAGGAGAGAACTCTCA	2628
Oy	397	TrpGLYArGLINProHisSerAsnArgArpheSerSerArGArTrgaAspAspSerSer	416
Db	2629	-----AAAGAGACCATGAGAAC-----ACAGAGAAATGCATGAGGACGA	2667
Oy	417	GlusErspAspSerGlyserSerSerGluSerasp	428
Db	2668	GAAAGCCAGGGGCTGCTGCTGCACAGACCTGAGAT	2703
RESULT 2			
US-09-785-770A-14			
; Sequence 14, Application US/09785770A			
; Patent No. US20020103360A1			
GENERAL INFORMATION:			
APPLICANT: Pan, Yang			
TITLE OF INVENTION: A NOVEL PROTEIN RELATED TO MELANOMA-INHIBITING PROTEIN			
FILE REFERENCE: 07334-328001			
CURRENT APPLICATION NUMBER: US/09/785,770A			
PRIOR APPLICATION NUMBER: 2001-02-16			
PRIOR FILING DATE: 1999-09-01			
PRIOR APPLICATION NUMBER: US 09/145,056			
PRIOR FILING DATE: 1998-09-01			
NUMBER OF SEQ ID NOS: 24			
SOFTWARE: Patent In Ver. 4.0			
SEQ ID NO 14			
LENGTH: 8121			
TYPE: DNA			
ORGANISM: Homo sapiens			
FEATURE:			
NAME/KEY: 5'UTR			
LOCATION: (1)...(4)			
NAME/KEY: CDS			
LOCATION: (5)...(5725)			
NAME/KEY: 3'UTR			
LOCATION: (5726)...(8121)			
US-09-785-770A-14			
Alignment Scores:			
Pred. No.:	8.25e-05	Length:	8121
Score:	148.00	Matches:	106
Percent Similarity:	34.76%	Conservative:	65
Best Local Similarity:	21.54%	Mismatches:	209
Query Match:	6.49%	Indels:	112

DB: 10 Gaps: 24

US-09-700-696b-2 (1-430) x US-09-785-770A-14 (1-8121)

QY 1 Valaenlysglutyserilserasnlys-----GluasnThrHisasnly 16

DB 1376 GTAAACGCAAAACATCATTAAGAAAAGGAGGAGTTCAGAAATCCAGAGGGCC 1435

QY 17 Leu-----ArgmetSerIleTyrPro 23

DB 1436 CTGGTCAAGATGAGACAGAAATTAAGAGTGAATCAAGAGCCATGACGTCCACACT 1495

QY 24 LysSerThrGlyAsnLys-----GlyPheGluaspGlyaspaspalaile 38

DB 1496 TCTGTTCACAGCAATTAACCTCACTATGCCACCTGCTGTAAGGCTAAAGACATTA 1555

QY 39 SerLysLeuHisAspGlnGlu-----TyrGlyAlaIleLeu-----IleArgAsn 54

DB 1556 AAATCAGCTTATGATGATACAGAAATGACCTAAAGAGAGCAGCTATTCATCTCAAAA 1615

QY 55 AsnMetGlnHisIleMetGlyProValThrAlaIleLysLeuGlyGluAsnLys 74

DB 1616 GGAATGCTCCAGCAAGAAAAGCTGAGAGAGATTTGGAAAGTGGCTCAGAGAGTGA 1675

QY 75 -----GluasnThrProArgAsnValLeuasn-----IleIleProAlaSermet 89

DB 1676 TCTCCACAGAAAGCTGACAGCAATCAATGAATGACAGAAAGATTCACAGATCCCTG 1735

QY 90 AsnThrAlaLysAlaHisSerLysAspLysLysProGlnArgAspSerGlnAlaGln 109

DB 1736 GGATGTCACACATCATGAGGAGATGACCCCTAACCCATCCAGAGACATGTGGAGGA 1795

QY 110 LysSerProValLysSerLysSerThrHis-----ArgIleGlnHisAsnIleAspTyr 127

DB 1796 GACGCTTGGTAAATGGGGCCAAACTGCACACCTTTCATGAGCATCAACCTGAGAA 1855

QY 128 LeuLysHis-----LeuSerLysValLysLysIleProSerAspPheGluGlySerGly 145

DB 1856 TTGAAAGAGAAATTAGTCTTAAACTCAAAACCAACTGA---TTCCTCTCCAGAT 1912

QY 146 TyrThrAspLeuGlnGluArgGlyAspAsnAspIleSerProPheSerGlyAspGlyGln 165

DB 1913 GAGATGATTGTTCCACAGAGAACTGAAGACGAGTT---CCCATCTGGGAAAGAAATCTT 1969

QY 166 ProPhe-----LysAspIleProGlyLysGlyGlnAlaThrGlyProAspLeuGlu 182

DB 1970 CCTGGCACAGAAAGAGATGTG-----GCTGCCACACCCAGTAAGCAATG 2017

QY 183 GlyLysAspIleGlnThrGlyPheAlaGlyProSerGluAlaGlySerThrHisLeuAsp 202

DB 2018 AGTGAGAGATTAAG-----CTCTGAGGGAGAAAGCCAAAGAGAGACTCTTGAT 2068

QY 203 ThrLysLysProGlyTyrAsnGluIleProGluArgGluAsnGlyLysAsnThrIle 222

DB 2069 -----GAAAGATTGTTTTCATCAACAGCAATGACAGGACAGAGTA 2110

QY 223 GlyThrArgAspGluThr-----AlaLysGlu 231

DB 2111 GGACACAGACCAAACTACACAGCAGAGAGACCAAGCTTCTTCTTAAGTAAGAGG 2170

QY 232 AlaAspAlaValAspValSerLeuValGlyLysAsnAspIleMetGlySerThrAsn 251

DB 2171 GATGATTATCCCTGCAAGACTAGAGATGAAGAGATGAATGAATGCAAAAGGGCT 2230

QY 252 PheLysGluLeuProGlyArgGlyLysAsnArgValAspAlaGlySerGln----- 268

DB 2231 AAAAGAAAAAACCCGGGAGATCAGGGCAGGCTGTTGATGTTAATCTCAAGTCCCTGAC 2290

QY 265 AsnAlaHisGlyGlyLysValGluPheHisTyrProProAlaProSerLysGlyLysArg 288

DB 2291 AGAGCATTTTAAAGGACCATTCAT-----CCAGATCCCAAGAAATTCAGAAAGC 2338

QY 289 LysGluGlySerSerAspAlaAlaGlySerThrAsnTyrAsnGluIleProLysAsnGly 308

DB 2339 AAGCAAGAAACTGATGATTTGGATAGTGAAAAACAAGTAC----- 2383

QY 309 LysGlySerThrArgLysGlyValAspHisSerAsnArgAsnGlnAlaThrLeuAsnGlu 328

DB 2384 -----ACTGCTGCCAAAGGGGTCAACACAGAGGAGGAGCAACCAATCAATGCTGCA 2437

QY 329 LysGlnArg-----PheProSerLysGlyLysSer 338

DB 2438 AAAGAACGCCCTCTGGCAGTAAGAAAGACAGAGACCATTTGAAACAGTACCTTTCT 2497

QY 339 GlnGlyLeuProIleProSerArgLysLeuAspAsnGluIleLysAsnGluMetAspSer 358

DB 2498 GACAGCATAAATTTCAAGACTCCAGAAATTAAGTGAAGTCTTTCAGAAATTAAGATTCGTGAT 2557

QY 359 Phe-----AsnGlyProSerHisGluAsnIleIleThrHisGlyArgLysTyrHisTyr 376

DB 2558 TATCTGAAGACAGCAACCTGAGAAACATCTGAAGACCTCAGGGCTT----- 2605

QY 377 ValProHisArgGlnAsnAsnSerThrArgAsnLysGlyMetProGlnGlyLysSer 396

DB 2606 -----GCAGGGGACCTGAGAGGAGAACTGCA 2632

QY 397 TrpGlyArgGlnProHisSerAsnArgArgPheSerSerArgArgAspAspSerSer 416

DB 2633 -----AAGAGAGACCATGAGAAC-----ACAGAGAGTACATGGGCACA 2671

QY 417 GluSerSerAspSerGlySerSerSerGluSerAsp 428

DB 2672 GAAAGCCAGGGGTCTGCTGTCAGAAACCTGAAGAT 2707

RESULT 3

US-10-153-273-1

Sequence 1, Application US/10153273

Patent No. US20020169305A1

GENERAL INFORMATION:

APPLICANT: Sim, Kim L.

Chilnis, Chetan

Miller, Louis H.

Peterson, David S.

Su, Xin-Zhuan

Wellens, Thomas E.

TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS

NUMBER OF SEQUENCES: 37

CORRESPONDENCE ADDRESS:

ADDRESSEE: Knobbe Martens Olson & Bear

STREET: 620 Newport Center Drive 16th Floor

CITY: Newport Beach

STATE: California

COUNTRY: US

ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/153, 273

FILING DATE: 21-May-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/210, 288

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Fuller, Michael

REGISTRATION NUMBER: 36,516

REFERENCE/DOCKET NUMBER: NIH121.1FWDV1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 235-8550

TELEFAX: (619) 235-0176

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:


```

:      LENGTH: 4084 base pairs
:      TYPE: nucleic acid
:      STRANDEDNESS: single
:      TOPOLOGY: linear
:      MOLECULE TYPE: DNA (genomic)
:      HYPOTHETICAL: NO
:      ORIGINAL SOURCE:
:      ORGANISM: Plasmodium vivax
:      SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-153-273-1

```

Alignment Scores:	
Pred. No.:	0.000768
Score:	134.50
Percent Similarity:	36.42%
Best local Similarity:	19.96%
Query Match:	5.90%
DB:	9
Length:	408
Matches:	97
Conservative:	80
Mismatches:	220
Indels:	89
Gaps:	21

US-09-700-696B-2 (1-430) x US-10-153-273-1 (1-4084)

OY	2	AsnysglutylserIleSerAlnslysglunThrHisasnLytleuAdgmetSerIle	21
		: ::	
Db	1724	AATGAATTCAATGACTATAAAAAAGCGACAAAGAAGTTCCAGACGGCATTCGTACCTC	1783
OY	22	TyrProLysSerThrglyAsnLysglYrhegluaSprglyAspAspAlalIeSerLysLeu	41
		: ::	
Db	1784	TAT-----GANTTCTTAACAAG---1801	
OY	42	HisapcInglucgtYrrglYalAlaleuIleldgAsnAsnmeGlInHisIlemeGly	61
		::: :: :::	
Db	1802	--GAGTTAGATGAAATTAAACGAGTGCGCTTTTGAAAGAAATTAACAAACGTGATGT	1858
OY	62	ProvalThralIeLysLeuengLy--GluglyAsnLysglYasThrProtyasn	80
		:::	
Db	1859	GCAATATATGAGTTATGCCCTTTCTCCGTGAAGAGCGCTAAAATAAATCTCACAGAGTT	1918
OY	81	ValLeuasnlIelPerolaSerMetAsnTYrrAlalALyalaHisSerLysAspLys	100
		:: :::	
Db	1919	GTGCACAAATGTGGACAAATGCGTAAATCTCAGGCC---ACCANATCAAATCCGATATGT	1975
OY	101	LysProGlnArgaspSerGlnAlngLysSerProValLysSerLysSerThHisAlg	120
		::: : :::	
Db	1976	CAGCTGTGATAGTACTAGTAAAGCGGAAAGGTTCCA---GGAGATTCTACGAT---	2026
OY	121	IleGlnHisAsnIleaspTYrLeuLysHisLysSerLysVal---LysLysIleProSer	139
		: :::	
Db	2027	----GGAAATGTAACTAGTGGCCCAAGATAGTCTTACACAGATGAAGCTGTATCAGGG	2080
OY	140	AspPhegluglySerglYTYrThrAspLeuGln-----GluArgglyAspAsnAsp	156
		:: : :::	
Db	2081	GATGGTCAAAATGTGAAMTCAAGACACCTGCGAAGAACGATGTACACGAAAGTATGTCC	2140
OY	157	IleSerProPheSerGlyAspBglyGlnProPhelYAspRleProglYlysGlyLu---	175
		: : :::	
Db	2141	GAAAGTGTAAAGTCTAAAAAATGTTGATCCGACGAATCTGTAACTAAAGAAAGTAGCAC	2200
OY	176	-----AlatHngLYrProAspLeuGlnGLYAspRIleSglnThrclyrPheAlagly	192
		:::	
Db	2201	ACTGCACAGCGTTACAGATTTGCCGGAAGCTGGAAAGGA---AACTTAGCGCATCAAT	2257
OY	193	ProSerGlnAlagLuserrThrlHisLeuAspThrlLysLysProgly---	207
		::: :::	
Db	2258	AGTCGACCTTTCGATGCACG---GTTGAACAAATAGCCACAGTGATGATACTGTGAAC	2314
OY	208	-----TyraSngLule	211
		:::	
Db	2315	AGTCATCTATACCTGTAGTAGTGSTGA AAAACCCATTTGTAACCCCTTATATGTTTTG	2374
OY	212	ProGluArgglungLynsglyYasn-----Thr	221
		: :::	
Db	2375	AGGCATTCGAAAGACAAATATGATAGACGATAGACCTCGGAAATCATATGGCGAATCOTGAT	2434

QY	222	IIeGIyThArGAspGclUThPrAlaLySGlUAlaSPAlaValaSPAlaSerLeuValGlu	241
Db	2435	TCAATAGTAAAGGTGACGAGGGAAGGGCCAAAGATAAATGATATGGCGAAGGCTACTAA	2494
QY	242	GIYSerAsnAPILMeTcIYSerThAsnPheLySGluLeuProGlyArGluGlyAsn	261
Db	2495	GATAGTACTGTAATACTTCAGATGTCACGACG-----TCTGCTCAAGGTGTAT	2539
QY	262	ArgValAspAlaGlySerGlnAsnAlaHisGlnGlyLyValGluPheHisTryProPro	281
Db	2540	ACTACTGTCGACAGTTGATAGGAAATTAAAGGTGCTCTGAGATAGCGATAAACT	2599
QY	282	AlaProSerLySGluLyAsArgLySGluGlySerSer-----AspAlaAlaGlySer	298
Db	2600	GTAGAGACTAAAGATGAGAGGGGAGAGATACTCTGCAATAAAGATGACGACGACTGTA	2659
QY	299	ThraSNTrYAsnGluIlleProLySAsnGlyLySGlySerThr-----ArgLySGly	315
Db	2660	GTTGCTGAGTGATTAACATTCGTGAGACACGCGCTGCTGTAGCACATAATGATATAGTAA	2719
QY	316	ValAspHisSerAsnAPcAsnGlnAlaThrLeuAsnGluLySGlnArg-----PhePro	333
Db	2720	AATGACACGGAAAGAACGGGGCCTTACCCCTGACAGTAAACAACTAGATGACGACT	2779
QY	334	SerLySGlyLySerGlnGlyLeuProIlleProSerArgGly-----LeuAspAsnGlu	351
Db	2780	CGCGTAACTAAACCGAAAGTTTAGAATTCACAAACAGAAAGTGAAGTAACTACTATGAT	2839
QY	352	IIeLySAsnGluMeLAspSerPheAsnGlyProSerHisGluAsnIlleIleThrHisGly	371
Db	2840	ACAACACTACAGTTTAGAATTAATAATGAGAGAAAGAAAGAAAGATTTACAAAGCATGAT	2899
QY	372	ArgLySTYrHISTryValProHisArgGlnAsnAsnSerThArg-----	386
Db	2900	TTTAAAGATATGATACCGCGCAATGAAACCAAAATTCGTATCAACTACAGATGCAGAA	2959
QY	387	-----AsnLySGlyMeLProGlnGlyLySGlySerTrpGlyArgGlnProHisSer	403
Db	2960	GCACATGACAGGATAGCATCAAAAATGATAAACGACA-----AGAGAAAGCATATG	3013
QY	404	AsnArg-----ArgPheSerSerArgArgArgAspAspSerSerGluSerSerAspSerGly	422
Db	3014	AATTAAGATACTTTTACGAAATAATCAAAATAGTCCATTTAAATAGTAATAATAATTG	3073
QY	423	SetSerSerGluSerAsp	428
Db	3074	AGTAATGCAAAATTAGAT	3091
RESULT 4			
US-09-864-761-4620/c			
Sequence 4620, Application US/09864761			
Patent No. US20020048765A1			
GENERAL INFORMATION:			
APPLICANT: Penn, Sharon G.			
APPLICANT: Rank, David R.			
APPLICANT: Hanzel, David K.			
APPLICANT: Chen, Wensheng			
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR			
FILE REFERENCE: Aeonica-X-1			
CURRENT APPLICATION NUMBER: US/09/864,761			
CURRENT FILING DATE: 2001-05-23			
PRIOR APPLICATION NUMBER: US 60/180,312			
PRIOR FILING DATE: 2000-02-04			
PRIOR APPLICATION NUMBER: US 60/207,456			
PRIOR FILING DATE: 2000-05-26			
PRIOR APPLICATION NUMBER: US 09/632,366			
PRIOR FILING DATE: 2000-08-03			
PRIOR APPLICATION NUMBER: GB 24263.6			
PRIOR FILING DATE: 2000-10-04			
PRIOR APPLICATION NUMBER: US 60/236,359			
PRIOR FILING DATE: 2000-09-27			
PRIOR APPLICATION NUMBER: PCT/US01/00666			

```

PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
SEQ ID NO 4620
LENGTH: 1960
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC006059.3
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.1
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.2
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.2
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.8
US-09-864-761-4620

Alignment Scores:
Pred. No.: 0.000368
Score: 133.50 Length: 1960
Percent Similarity: 36.08% Matches: 88
Best Local Similarity: 18.57% Conservative: 83
Query Match: 5.86% Mismatches: 206
DB: Indels: 97
Gaps: 17

US-09-700-696b-2 (1-430) x US-09-864-761-4620 (1-1960)
QY 2 Asnlysglutyserilserasnllysglunthrhisainglyleuargmetserile 21
DB 1916 AGTAACATCAAGTCAGATCTGATCCCAAGCTGGCCGCA---AAGAGAGCATCAAAA 1860
QY 22 TyrProlyserthnglyasnllysglyphgclunspgelyspapalailserlyleu 41
DB 1859 TCACCAAGAAAACACGCTCTCACTTAAGTGAAGTAACCAAGTAAACAGAACCTTTA 1800
QY 42 Hisapglnglglutyrlyalalaileuilearvasnsmetglnhilemetglj 61
DB 1799 AGA-----GCAACATGCGCACAAATGAAATGTAGTAGTCAAA 1761
QY 62 Provalthralailyleuleuglyglunlunslunthrrproarvasnval 81
DB 1760 CCACTTAGTACA-----GAAATATATCTCT----- 1737
QY 82 leuansnleileproalasermetasnryalalysalaihiserlyspylslys 101
DB 82 leuansnleileproalasermetasnryalalysalaihiserlyspylslys 101
DB 1736 -----GTATACCACTGAGTGCACAGTCCCCCTTCAGATGAGAGCCGTGACAGAAA 1683
QY 102 Proglinargaspserglunlagnlyserprovalyaserlyserthrhisaingly 121
DB 1682 CCTGGAGAGCCCTTATGAGCAATTCAGAAATGCAAAAGCTTAAACCAACCTTTGCTA 1623
QY 122 GlnHisasnileasptyrleuylshisleserlyvalylslytleproserasphe 141
DB 1622 CCCATCCAAAGCCCTTAC-----AGTTAGCAAAATTTAAAGACACTGGTAGC----- 1575
QY 142 Gluglyserglytyrthraspheuglunlagnlyaspasaspilserprohser 161
DB 1574 ---TCATCATCTTACCATTAAGAGAAAATAATGGGAAGAGTACAGACTTATTTCA 1518
QY 162 Glyaspglglinprophelyaspilproglylysglylualathrglyproasphe 181
DB 1517 -----AAATACAGTGT-----AGACCTTCAGAAAGCTCACCAAGCTCA 1479
QY 182 Gluglylyaspilleglthrghlyphealaglproserglunlagnlyserthrhisleu 201
DB 1478 AGGAGCAGATCTTACAGAGTATCTTATCCAGATCATATPACAGATTCAGTCTTA 1419
QY 202 Aspthrlyspgelytyrthrasnglulleproglunlagnlyaspilserthrh 221
DB 1418 ---GCTAGTTCATCATTCAGAGTCTAGCTCTCCATCATCTATCATTCAGAAATTA 1362
QY 222 Ileglythrargaspilglthrghlyphealaglproserglunlagnlyaspilserthrh 241
DB 1361 TACAGTATCATTCACAGTCTAGTATCATCTTATCATCTTATTCAGTATGAT 1302
QY 242 Gly----- 242
DB 1301 GGAAGGCCAGGTAGAGAGAGACTTAGTACAGTGGAAAAAATAGCGTTTACATTA 1242
QY 243 -----Seranaaspillemetglyserthrasphelysglueuproglyargglu--- 239
DB 1241 AAGCATGACACACACTCTGAAAGACACTTCACGTAAATATGTCAAAGGTAAAGACAG 1182
QY 260 -----Glyasnargyalasplaglserglnasn 269
DB 1181 TCTTCATGTGTGAGAAAGTATAGGAGAGCATCATCTTATGATAT---TCTTACAGC 1125
QY 270 AlaHisngllylyvaligluiphehistryrproalaproserlysglyllyas 289
DB 1124 AGTAGAGATGACAGTGTTCAG-----GCCACAGTACAGCCAGGAAAAAGAGAG 1074
QY 290 Gluglyserasplala-----GluSer 298
DB 1073 CAGGGCCAAATGGAAGAACATATTAACAGAAAAAAGAGAGTGAAGAAAAATCC 1014
QY 299 ThrAsnlyrasnglulleprolyasnglylysertharglysglyllyasph 318
DB 1013 AAGTCTGACGCGAATGCCCTCATTCAAAAAAGAACTTGAAGAGATCTTCTGAT 954
QY 319 Serasnargasnlnalathrhleuasnngllylysglnargpberpserlysglylyser 338
DB 953 CACCTTAGAAGATGGCAGTAAGCCAAAGAAAGAAATTAATGCTGGTAGTAATGGACTCT 894
QY 339 GlnGlyleuproileproserargelyleuaspaasnnglullelyasnglu----- 355
DB 893 GAGTCA-----AATTCAGACGAGATGCTCAAAAAACAGTAAATAATGACATCCATCA 840
QY 356 ---Metaspserpheasnnglyproserhnglunlilethrhngllyarglytyr 374
DB 839 TCCTGTCAGCAAGAAAGAGTGAAGGACCAATCCGATTTGTGAATTCAGAGCTTAGTGAATTT 780
QY 375 Hislyvalprohishargylasnasnserthrhargasnlysglymetplogllyly 720
DB 779 CACATCAAGTCAAAACCAACCAAGTGCACCAAAATACTTCACTCCGATGATAT 720
QY 395 glyserTPgllyarglprohishiserasnarghrpbeserSerargaglyaspasp 414
DB 719 GGTCTGTGAAA-----TCAGCAAAACAGCCGACATCA 687

```



```

APPLICANT: Macina, Roberto
APPLICANT: Recipon, Hervé
APPLICANT: Cefferey, Robert
APPLICANT: Sun, Yonngming
APPLICANT: Liu, Chenghua
TITLE OF INVENTION: Compositions and Methods Relating to Ovary Specific Genes and Pro
FILE REFERENCE: DEX-0277
CURRENT APPLICATION NUMBER: US/10/001,835
PRIOR FILING DATE: 2001-11-20
PRIOR APPLICATION NUMBER: 60/249,997
NUMBER OF SEQ ID NOS: 228
SOFTWARE: PatentIn version 3.1
SEQ ID NO 66
LENGTH: 1450
TYPE: DNA
ORGANISM: Homo sapiens
US-10-001-835-66

```

```

Alignment Scores:
Pred. No.: 0.00497          Length: 1450
Score: 130.50             Matches: 90
Percent Similarity: 35.60%  Conservative: 46
Best Local Similarity: 23.56%  Mismatches: 121
Query Match: 5.73%         Indels: 129
DB: 9                      Gaps: 19

```

US-09-700-696b-2 (1-430) x US-10-001-835-66 (1-1450)

```

QY 79 ArgAsnValLeuAsnIleIleProAla-----SerMetAsnTyrAlaValAlaHisSer 96
DB 411 CGACAGCTGTCTAACAACACTGTGGGATACCTACTGTGGACACACTGGGGTTCACAAAG 470
QY 97 LysAspLysLysLysProGlnArgAspSerGlnAlaGlnLysSerProValLysSerLys 116
DB 471 AGCAAGCAAAACAAACACACAC-----ACAGCAAAAGAAAGACAGAAACAAACAA 524
QY 117 SerThrIleArgIleGlnHisAsnIleAspTyrLeuLys-----HisLeuSerLysVal 134
DB 525 AAGACGGAGAGACGAGGAGAGAT---ACACACACAAACACGGGCACACAAACAAAGAG 581
QY 135 LysLysIleProSerAspPheGlnLysSerGlyTyrThrAspLeuGlnLysArgLys 154
DB 582 CAGAGGTACACACACGAC-----AGTGGGTGTACCAAAACACAGAG----- 623
QY 155 AsnAspIleSerProPheSerLysAspLysGlnProPheLysAspIleProGlyLysGly 174
DB 623 ----- 623
QY 175 GlnAlaThrGlyProAspLeuGlnLysAspIleGlnThrGlyPheAlaGlyProSer 194
DB 624 -----ACAGAAAGAGCGAGCCACCC 644
QY 195 GlnAlaGlnSerThrHisLeuAspThrLysLysProGlyTyrAsnGlnIleProGlnArg 214
DB 645 AACAAAGAAAT-CAACACACACCAACAAAGAAACCA-----CATCCACGTGGCGCCGC 697
QY 215 GlnGluAsnGly---GlyAsnThrIleGlyThrArgAspGlu----- 227
DB 698 GAGAGAGGGGGAGGAGGAGACGAGGAGGAGAGAGAAAGAAAGAGAGAGAGC 757
QY 228 ---ThrAlaLysGlnAlaAspAlaValAspValSerLeuValGlnGlySerAsnAspIle 246
DB 758 GCGGCGCAGAAAGAGAGAGAGAGAG-CAACGCGCAGAAAGAGAGAGAGAGAGAA--- 813
QY 247 MetGlySerThrAsnPheLysGlnLeuProGlyArgLysGlnLysAsnArgValAspAlaGly 266
DB 814 -----AAAAAGAGACCAAGGAGAGAGAGAGGCGCAGACAGCGAGC 855
QY 267 SerGlnAsnAlaHisGlnGlyLysValGlnPheHisTyrProProAlaProSerLysGln 286
DB 856 GAGGGGAGAGAAAC-----TCG 873

```

```

QY 287 LysArgLysGlnLysSerSerAspAlaIleGlnSerThrAsnTyrAsnGlnIleProLys 306
DB 874 AGGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 932
QY 307 -----AsnGlyLys---GlySerThrArgLys 314
DB 933 GAGGACACCAAGGCGAGGAGGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 992
QY 315 Gly-----ValAspHisSerAsnArgAsnGlnAlaThrLeuAsnGlnLysGlnArg 331
DB 993 GCGGACGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1052
QY 332 PheProSerLysGlyLysSerGlnLysLeuProIleProSerArgGlyLeuAspAsnLys 351
DB 1053 ACCGCA---AGAGAAAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 1109
QY 352 IleLysAsnGlnLeuMetAspSerPheAsnGlnLysProSerHisGlnAsnIleThrHisGly 371
DB 1110 GACAGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1135
QY 372 ArgLysTyrHisTyrValProHisArgGlnAsnAsnSerThrArgAsnLysGlyMetPro 391
DB 1136 -----CACAGGACGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 1174
QY 392 GlnGlyLysGlySerTyrGlyArgGlnProHis-----SerAsnArgArgPheSerSer 409
DB 1175 GCAGGCACA-----GCACGACAGAGACACAAACAAAGACAGAGAGAGAGAGAGAG 1225
QY 410 ArgArgArgAspAspSerSer-----GluSerSerAsp 420
DB 1226 AAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1285
QY 421 SerGly 422
DB 1286 GAAAGCA 1291

```

```

RESULT 7
US-09-962-436-563
Sequence 563, Application US/09962436
Patent No. US2002008101A1
GENERAL INFORMATION:
APPLICANT: Soppet, Daniel
TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Sign
FILE REFERENCE: 689280-75
CURRENT APPLICATION NUMBER: US/09/962,436
PRIOR FILING DATE: 2001-09-25
PRIOR APPLICATION NUMBER: US/60/235,082
PRIOR FILING DATE: 2000-09-25
PRIOR APPLICATION NUMBER: US/60/234,924
NUMBER OF SEQ ID NOS: 568
SOFTWARE: PatentIn version 3.0
SEQ ID NO 563
LENGTH: 2454
TYPE: DNA
ORGANISM: Homo sapiens
US-09-962-436-563

```

```

Alignment Scores:
Pred. No.: 0.00125          Length: 2454
Score: 129.50             Matches: 95
Percent Similarity: 34.52%  Conservative: 60
Best Local Similarity: 21.16%  Mismatches: 171
Query Match: 5.68%         Indels: 123
DB: 10                      Gaps: 22

```

US-09-700-696b-2 (1-430) x US-09-962-436-563 (1-2454)

```

QY 54 AsnAsnMetGlnHisIleMetGlyProValThr-----AlaIleLysLeuLeuGlnGly 71
DB 182 GATTAACAGAACACACAAATGAGATGCTGCTCATTCATTCAGTCTCTCAAT 241

```

```

Oy 72 Glu---AsnLysGluAsnThrProArgAsnValLeuAsnIleIleProAlaSerMetAsn 90
Db 242 GCCTTCTCGAAGTCCAGCGCTCCA-----CCCATCACTCTAGTCCGCCCA 289
Oy 91 TyrAlaLysAlaHisSerLysAsp---LysLysLysProGlnArgAspSerGlnAlaGln 109
Db 290 GTCCTGAGACGAGTAGAAGAACGTCMAAGAACAGACACACTGAAATGAAGACACA 349
Oy 110 LysSerProValLys-----SerLysSerThr 118
Db 350 AAGTTGAAGTAAGATGTTAGAGACCCAGCTGATGCTCGAAGCCCAAGCTCTCC 409
Oy 119 HisArgIleGlnHisAsnIleAspTyrLeuLysHisLeuSerLysValLysLysIlePro 138
Db 410 AGCAGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 469
Oy 139 SerAsp-----PheGluLysSerGlyTyrThrAspLeuGlnLysGlyAspAsn--- 155
Db 470 ACAGAGAAATGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 523
Oy 156 AspIleSerProPheSerGlyAspGlyGlnProPheLysAspIleProGlyLysGlyGlu 175
Db 524 CAGTGGACCTCTATCCCTCCGACAGCCAACTCTGAAAGAGAGAGAGAGAGAGAG 583
Oy 176 AlaThrGly-----ProAspLeuGluLysAspIleGlnThrGlyPhe 190
Db 584 GAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 643
Oy 191 AlaGlyProSerGluAlaGluSerThrHisLeuAspThrLysLysProGlyTyrAsnGlu 210
Db 644 CGAGGGAGAGATAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 697
Oy 211 -----IleProGluArgGluGluAsnLysGlyAsnThrIleGlyThrArgAspGlu 227
Db 698 AACGCTTTCTCAATGAAGAAAGAGAG-----GCTTCA 730
Oy 228 ThrAlaLysGluAlaAspAlaValAspAlaSerLeuValGluLysSerAsnAspIleMet 247
Db 731 GCTATATAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 790
Oy 248 GlySerThrAsnPheLysGluLeuProGlyArgGluLysAsnArgValAspAlaGlySer 267
Db 791 GAGAAAGACACATAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 847
Oy 268 GlnAsnAlaHisGlnGlyLysValGluPheHisTyrProProAlaProSerLys----- 285
Db 848 CAGGAGAAATCAC-----CCACAGAGACTCTAAAGCCCA 880
Oy 286 -----GluLysArgLysGluLysSerSerAspAlaIle----- 296
Db 881 CCCCAGAACCCAGAGAAATCTGAGAAAGGTGAGAAAGATGACCTGAGGTGACAA 940
Oy 297 -----GluSerThrAsnTyr 301
Db 941 CGAGCAGCAGAGCCAGACACACACAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1000
Oy 302 AsnGluIleProLysAsnGlyLysGlySerThrArgLysGlyValAspHisSerAsnArg 321
Db 1001 GGGAGTCTTCTCTGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1060
Oy 322 AsnGlnAlaThrLeuAsnGluLys-----Gln 330
Db 1061 AGCATGGCCAGTTTAGGGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1120
Oy 331 ArgPheProSerLysGlyLysSer-----GlnGlyLeuProIleProSer--- 345
Db 1121 CAAGACCTGAATATGAGAGAAATTAAGAGGTTATCCAGAGCCTCCAGAGAGAG 1180
Oy 346 -----ArgGlyLeu 348
Db 1181 CTGAGATGGAGAGCCTATAGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1240
Oy 349 AspAsnGluIleLysAsnGluMetAspSerPheAsnGlyProSerHisGlu---AsnIle 367

```

```

Db 1241 AGTGAGAGAGTGGATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1300
Oy 368 IleThrHisGlyArgLysTyrHisTyrValProHisArgIleAsnAsnSerThrArgAsn 387
Db 1301 ATGGACACTGGA-----TATGGTGAAGAAAGTGAAGAGAGAGAGAGAG 405
Oy 388 LysGlyMetProGlnGlyLysGlySerThrP-----GlyArgGlnProHisSerAsnArg 423
Db 1337 AGGGCCCTTGAGCCGGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 440
Oy 406 ArgPheSerSerArgArgArgAspAsp 414
Db 1397 TATTTCATGCTGACACAGAGAGAGAG 1423

RESULT 8
US-09-880-107-3423
Sequence 3423, Application US/09880107
Patent No. US20020142981A1
GENERAL INFORMATION:
APPLICANT: Horne, Darci T.
APPLICANT: Vockley, Joseph G.
APPLICANT: Scherf, Uwe
APPLICANT: Gene Logic, Inc.
TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
FILE REFERENCE: 44921-5028-WO
CURRENT APPLICATION NUMBER: US/09/880,107
CURRENT FILING DATE: 2001-06-14
PRIOR APPLICATION NUMBER: US 60/211,379
PRIOR FILING DATE: 2000-06-14
PRIOR APPLICATION NUMBER: US 60/237,054
PRIOR FILING DATE: 2000-10-02
NUMBER OF SEQ ID NOS: 3950
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 3423
LENGTH: 4840
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: Genbank Accession No. US20020142981A1 U76366
US-09-880-107-3423

Alignment Scores:
Pred. No.: 0 00488 Length: 4840
Score: 127.50 Matches: 99
Percent Similarity: 33.05% Conservative: 54
Best Local Similarity: 21.38% Mismatches: 184
Query Match: 5.59% Indels: 128
Gaps: 19
DB:

US-09-700-696b-2 (1-430) x US-09-880-107-3423 (1-4840)
Oy 36 AsnAlaIleSerLysLysHisAspGlnGluTyrGlyAlaAlaLeuIleArgAsnAsn 55
Db 313 GACCCATACAGACCTCGAGAGAGCTGGAAGAGAGAGAGAGAGAGAGAGAGAGAG 371
Oy 56 MetGlnHisIleMetGlyProVal-ThrAlaIleLysLeuLeuGlyLysGluLysGln 75
Db 372 CAAGCCACCCCAAGACATAGCATCTACCAACTCTCAGTCTCGGGGCGGAGAG 423
Oy 75 uAsnThrProArgAsnValLeuAsnIleIleProAlaSerMetAsnTyrAlaLysAla 95
Db 424 -----TTGCCATCAAGCATG----- 438
Oy 95 sSerLysAspLysLysLysProGlnArgAspSerGlnAlaGln----- 109
Db 439 -----AAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 494
Oy 110 -----LysSerProValLysSe 115
Db 495 ACACCCCTGCACTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 554
Oy 115 rLysSerThrHisArgIleGlnHisAsnIleAspTyrLeuLysHisLeuSerLysVal 135

```

```

Db      555 AGCAGAGCCC-----TCAGCAAAATACGTGGTCTCAGAAACTGAGAGGAGG 605
Qy      135 slysileproser-----
Db      606 CAGGCTCCCGGCTTTGGAGCTGCTGCCAGCCCTGGATGGTGTACAGGGCCAGGCCGA 665
Qy      140 -----AspHegluglyserglytyrthrAspleuglInuArgglyAs 154
Db      666 CAGCTCCAGCAGACACCTCCAGCTCCAGATGATGACACAGCTGAGTAAAGGCTTC 725
Qy      154 pAsnAspIleSerProPheSerSerglyAspGlyGlnProPheLysAspIleProglyLysgl 174
Db      726 TGAATAATTTCTCCAGGTCAGAGCTGCTCAGCCCTCCAGAGGGAGCCCTGGAAAG 785
Qy      174 ygluAlaThrGlyProAspLeuGluGlyLysAspIleGlnThrGlyPheAlaGlyProse 194
Db      786 GCGTACCCAGCACCC-----CCTGG 806
Qy      194 rgluAlaGluSerThrHisLeuAspThrLys-----LysProglytyrAsnGluIlePr 212
Db      807 GAAGCAGGGGCTGTACCTCCAGACCAAGCAGAGGAAAGCA-----GA 851
Qy      212 ogluArgGluGluAsnGlyLysAsnThrIleGlyThrArgAspGluThrAlaLysGluAl 232
Db      852 GAGAGACTCAGACAGCAGCAGAGAGATCATCTGACAGTGAAGAGAGAGCCCACTGC 911
Qy      232 AspAlaValaAspValaSerLeuValaGluGlySerAsnAspIleMetGlySerThrAsn 252
Db      912 CAAGGCCCTCTTCAGGCGGAAAGCCTCAGGAAACCTCTCAGTGGAGGAGCTCCCA-- 969
Qy      252 elysGluLeuProGlyArgGluGlyLysAsnArgValaAspAlaGlySerGlnAsnAlaHisgl 272
Db      970 -----GCCCTCGCAAGAGATCCCCCAG-----AAAGAGAGTCCCCAGCGCCCC 1016
Qy      272 ngluLysValaGluPheHisIleTyProProAlaProserLysGluLysArgGluGlySe 292
Db      1017 TGGGAGACAGGGGCTCGCAAGTTCGCAAGCCCAAGCGGGGAAAGCGGAGAGACTCCGA 1076
Qy      292 rSerAspAlaAlaGluSerThrAsnTyArgAsnGluIleProLysAsnGlyLysGlySerTh 312
Db      1077 GAGCAGACGAGGAGATCGAGATGAGAGAGGAGCGCCCTGTCAGGGGAAAGCCTTCAGG 1136
Qy      312 rArgLysGlyValaAspHisSerAsnArgAsnGlnAlaThrLeuAsnGluLysGlnArg 332
Db      1137 GAAG-----GCCCCAGGTCAGAGCCGCTCGGCCCTGCCAAGAGAGTC 1181
Qy      332 eProSerLysGlyLysSerGlnGlyLeuProIleProserArg-----G 347
Db      1182 CCCGAGAAAGGGGCTGCC-----CCAGCACTCTCTTGAAAGAAAGGGCTCGCAGC 1232
Qy      347 yLeuAspAsnGluIleLysAsnGluMetAspSerPheAsnGlyProSerHisGluAsn-I 367
Db      1233 CCCCCAGCTCCAGTGGGAAAGCAGAGAGAGACTCAAGAAAGCAGAGCAGAGATCAGA 1292
Qy      367 leIleThrHisGlyArgLysTyArgLysTyArgValaProHisArgGln----- 381
Db      1293 CAGTGAACA-----CAGAAAGCATCGSCAGCATGATCATCACTCAGGTGAAGCCCTT 1343
Qy      382 --AsnAsnSerThrArg--AsnLysGlyMetProGlnGlyLysGlySerThrIleArg 400
Db      1344 GGGGAAAGCCCCCAGGTGAAGAACTGCTTACCATGGGCAATGGGCTTGGGGAAG 1403
Qy      400 InPro-----HisSerAsnArg-----ArgPheSer 409
Db      1404 CGCGGGCCCACTGCGCAGTGGGAAAGTGGGGCTCGCAACCCCTCAGCCAGGTGGGAA 1463
Qy      409 eArgArgArgAspAspSerSerLysSerSerAspSerArgLysSerSerGlySerArg 429
Db      1464 GTGGAGAGAG-GAC-----TCAGAGAGCAGTAGTGTAGAGATCATCAGACAGAGTGTAG 1516
Qy      429 LysAsp 430

```

```

Db      1517 GAGAG 1521
RESULT 9
US-09-746-660A-63
; Sequence 63, Application US/09746660A
; Publication No. US20030049804A1
; GENERAL INFORMATION:
; APPLICANT: Pompejus, Markus
; APPLICANT: Krogger, Burkhard
; APPLICANT: Schroder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Habermeyer, Gregor
; APPLICANT: Kim, Jun-Won
; APPLICANT: Lee, Heung-Schick
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING
; FILE REFERENCE: BGI-121CP2
; CURRENT APPLICATION NUMBER: US/09/746, 660A
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 09/606740
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: 09/603124
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: 60/141031
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: 60/142101
; PRIOR FILING DATE: 1999-07-02
; PRIOR APPLICATION NUMBER: 60/148613
; PRIOR FILING DATE: 1999-08-12
; PRIOR APPLICATION NUMBER: 60/187970
; PRIOR FILING DATE: 2000-03-09
; PRIOR APPLICATION NUMBER: DE 19931420.9
; NUMBER OF SEQ ID NOS: 125
; SOFTWARE: PatentIn Vers. 2.0
; SEQ ID NO 63
; LENGTH: 1066
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (101)..(1066)
; OTHER INFORMATION: RXC00866
US-09-746-660A-63
Alignment Scores:
Pred. No.: 0.000748
Score: 127.00
Percent Similarity: 34.50%
Best Local Similarity: 24.02%
Query Match: 5.57%
DB: 9
Gaps: 7
US-09-700-696B-2 (1-430) x US-09-746-660A-63 (1-1066)
Qy      191 AlAGlyProSerGluAla---GluSerThrHisLeuAspThrLysLysProGlytyrAsn 209
Db      143 CGCGGGCCCAACAGACCTGTCAGGAAACCAATCTGATACC----- 184
Qy      210 GluIleProGluArgGluGluAsnGlyLysAsnThrIleGlyThrArgAspGluThrAla 229
Db      185 CCGTCTTTCAGGCACACAGATGCTTCCTGTACAGAGCGGCTGTAAGAGCTGAGAGCCGC 244
Qy      230 LysGluAlaAlaValaValaAspValaSerLeuValaGluGlySerAsnAspIleMetGlySer 249
Db      245 GGAAGACAGAAATCGGAGATCTCGCAAGGTGCTCAAGATCCCAAGATTTCTCAGGGTTCC 304
Qy      250 ThrAsnPheLysGluLeuProGlyArgGlu---GlyAsnArgValaAspAlaGlySerGln 268
Db      305 CAGAACGCTCAGAGTTCCAGAGAACCGGAGATCGGAAACCAACACCGACCTTCCAAC 364
Qy      269 AsnAlaHisGlnGlyLysValaGluPheHisIleTyProProAlaProserLysGluLysArg 288

```

```

Db 365 AACAAACCGTCGCGT-----GGTCGTGGA 388
QY 289 LysGluGlySerSerAspAlaIaGluSerThrAsnTyAsnGluIleProIysAsnGly 308
Db 389 CGTCGTGATCCGGAACCGCAATGAGGGCGGCAACACAGCGGTAAACGACCGT 448
QY 309 LysGlySerThrArgLysGlyValAspHisSerAsnArgAsn----- 322
Db 449 CAGGGCGGAACCGTCGCAACCGCGTGCGGAGCGGCAACCGTGTAAAGTCGATGAG 508
QY 323 GlnAlaThrLeuAsnGlyLysGlnArgPheProSerLysGlyLysSerGlnGlyLeuPro 342
Db 509 GGTCCGATCTGACCCGCGCTGCCAGAGCCACCAAGGCGGCAACCGGTCTGCGT 568
QY 343 IleProSerArgGlyLeuAspAsnGluIleLysAsnGluMetAspSerPheAsn----- 360
Db 569 ATTACGCACCTTGGTGCATTTCCGAAATCGTCGCAACATGACCGTGTGAGTACAC 628
QY 360 ----- 360
Db 629 AACGCTGCTCATCGTGAAGTGTGCTCTTCCCATCTTCAGGTGAGCGAGCGGTT 688
QY 361 -----GlyProSerHisGluAsn----- 366
Db 689 GACCTGATCTCTCTGACTTGGCCCAATTGAGAGATCACTGCACCGCGTCGATGATG 748
QY 367 IleIleThrHisGlyArgLysTyHis 375
Db 749 GTGGTTACTCAAGGACGACGAAAGACCAC 775

```

```

RESULT 10
US-09-738-626-2161
; Sequence 2161, Application US/09738626
; Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAKO
APPLICANT: SENO, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738, 626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 2161
LENGTH: 2154
TYPE: DNA
ORGANISM: Corynebacterium glutamicum
US-09-738-626-2161

```

```

Alignment Scores:
Pred. No.: 0.00189 Length: 2154
Score: 127.00 Matches: 55
Percent Similarity: 34.50% Conservative: 24
Best Local Similarity: 24.02% Mismatches: 88
Query Match: 5.57% Indels: 62
DB: 9 Gaps: 7

```

US-09-700-696b-2 (1-430) x US-09-738-626-2161 (1-2154)

```

QY 191 AlaGlyProSerGluAla---GluSerThrHisLeuAspThrLysLysProGlyTyAsn 209
Db 43 GCGGGCCCAACGAGAGCTGGTCAGGAAACCATCTGGATACC----- 84
QY 210 GluIleProGluArgGluGluAsnGlyLysThrIleGlyThrArgAspGluThrAla 229
Db 85 CCGTCCTTCAAGGACCAAGATGATCTCTCTTAACCAAGACCGCTGTAAAGCTGAGACCGCC 144
QY 230 LysGluAlaAspAlaValAspValSerLeuValGluGlySerAsnAspIleMetGlySer 249
Db 145 GGAAACGACAAATCGGATGCTGCGCAGAGTCTCAAGATCCCAAGATCTCAGAGTTCC 204
QY 250 ThrAspPheLysGluLeuPheArgLysGlu---GlyAsnArgValAspAlaGlySerGln 268
Db 205 CAGAACGCTCAAGGTCTCCAGAACCGGACGATCCGGAACCAACACCGCACCGCTTCCAC 264
QY 269 AsnAlaHisGlnGlyLysValGluPheHisTyProProAlaProSerLysGluLysArg 288
Db 265 AACAAACCGTCGCGGT-----GGTCGTGGA 288
QY 289 LysGluGlySerSerAspAlaIaGluSerThrAsnTyAsnGluIleProLysAsnGly 308
Db 289 CGTCGTGATCCGGAACCGCAATGAGGGCGGCAACACAGCGGTAAACGACCGT 348
QY 309 LysGlySerThrArgLysGlyValAspHisSerAsnArgAsn----- 322
Db 349 CAGGGCGGAACCGTCGCAACCGCGTGCGGAGCGGCAACCGTGTAAAGTCGATGAG 408
QY 323 GlnAlaThrLeuAsnGlyLysGlnArgPheProSerLysGlyLysSerGlnGlyLeuPro 342
Db 409 GGTCCGATCTGACCCGCGCTGCCAGAGCCACCAAGGCGGCAACCGGTCTGCGT 468
QY 343 IleProSerArgLysLeuAspAsnGluIleLysAsnGluMetAspSerPheAsn----- 360
Db 469 ATTACGCACCTTGGTGCATTTCCGAAATCGTCGCAACATGACCGTGTGAGTACAC 528
QY 360 ----- 360
Db 529 AACGCTGCTCATCGTGAAGTGTGCTCTTCCCATCTTCAGGTGAGCGAGCGGTT 588
QY 361 -----GlyProSerHisGluAsn----- 366
Db 589 GACCTGATCTCTCTGACTTGGCCCAATTGAGAGATCACTGCACCGCGTCGATGATG 648
QY 367 IleIleThrHisGlyArgLysTyHis 375
Db 649 GTGGTTACTCAAGGACGACGAAAGACCAC 675

```

```

RESULT 11
US-09-764-176-6
; Sequence 6, Application US/09764176
; Patent No. US2002012753A1
GENERAL INFORMATION:
APPLICANT: NOTEBOEN, Mathieu Hubertus Maria
APPLICANT: DANEN-VAN OORSCHOT, Astrid Adriana Anna Maria
APPLICANT: ROHN, Jennifer Leigh
APPLICANT: WEISS, Bertram
TITLE OF INVENTION: APOPTIN-ASSOCIATING PROTEIN
FILE REFERENCE: 472505
CURRENT APPLICATION NUMBER: US/09/764,176
CURRENT FILING DATE: 2001-01-17
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn version 3.0
SEQ ID NO 6
LENGTH: 5331
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
OTHER INFORMATION: Description of Sequence: nucleic acid sequence of AAP-2
NAME/KEY: misc.feature
LOCATION: (3851)..(3900)

```


b
2367 GAGGCATCAATTACAATCAGGATATCAAGTCGGAAAAGAGGCCCGATTCTCCCCCGG 2400

CURRENT APPLICATION NUMBER: US/09/864,761
 CURRENT FILING DATE: 2001-05-23
 PRIOR APPLICATION NUMBER: US 60/180,312
 PRIOR FILING DATE: 2000-02-04
 PRIOR APPLICATION NUMBER: US 60/207,456
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: US 09/632,366
 PRIOR FILING DATE: 2000-08-03
 PRIOR APPLICATION NUMBER: GB 24263.6
 PRIOR FILING DATE: 2000-10-04
 PRIOR APPLICATION NUMBER: US 60/236,359
 PRIOR FILING DATE: 2000-09-27
 PRIOR APPLICATION NUMBER: PCT/US01/00666
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00667
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00664
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00669
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00665
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00668
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00663
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00662
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00661

PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00670
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: US 60/234,687
 PRIOR FILING DATE: 2000-09-21
 PRIOR APPLICATION NUMBER: US 09/608,408
 PRIOR FILING DATE: 2000-06-30
 PRIOR APPLICATION NUMBER: US 09/774,203
 PRIOR FILING DATE: 2001-01-29
 NUMBER OF SEQ ID NOS: 49117
 SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
 SEQ ID NO 3788
 LENGTH: 1975
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE: MAP TO AC000015.2
 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.9
 OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.8
 OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.4
 OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.9
 OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.5
 OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.5
 OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9
 OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.6
 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.4
 US-09-864-761-3788

Alignment Scores:
 Pred. No.: 0.00239 Length: 1975
 Score: 125.50 Matches: 83
 Percent Similarity: 36.19% Conservative: 73
 Best Local Similarity: 19.26% Mismatches: 159
 Query Match: 5.51% Indels: 117
 Gaps: 17
 DB:

US-09-700-696B-2 (1-430) x US-09-864-761-3788 (1-1975)

QY 34 GLYASPAALAIleSerLysLeuHisAspGlnGluGlyTyrGlyAlaAlaLeuIleArg 53
 DB 1114 GGGAATGAAGATTCAATATATATCAAGACTCGAATAAT-----AACTTACTAATA 1064
 QY 54 AsnAsnMetGlnHisIleMetGlyProValThrAlaIleLysLeuGluGlnGluAsn 73
 DB 1063 AATTAATCTTCTGAAAC-----ACCAATTGGAAAAATCTTACAGAAAG 1019
 QY 74 LysGluAsnThrProArgAsnValLeuAsnIleIleProAlaSerMetAsnTyrAlaLys 93
 DB 1018 AATGAATCGCTGACCAACAT-----CTAGATCTACAGAGTTGCCTTAA 974
 QY 94 AlanHisSerLysAspLysLysLysProGlnArgAspSerGlnAlaGlnLysSerProVal 113
 DB 973 ACACACATTTGACACAGATTCAGAACATTTTAGTGAGACACAAATGAATGATACCTATG 914
 QY 114 LysSerLysSerThrHisArgIleGlnHisAsnIleAspTyrLeuLysHisLeuSerLys 133
 DB 913 GAGTGATTCATTTTCCAGTACCAAAATGA-----881
 QY 134 ValLysLysIleProSerAspPheGluGlySerGlyTyrThrAspLeuGlnGluArgGly 153
 DB 880 -----TCTGAAGTTGAACCATCTGTAAATGCTATCTTAACAAATGAAT 836
 QY 154 AspAsnAspIleSerProPheSerGlyAspGlyGlnProPheLysAspIleProGlyLys 173
 DB 835 GAAATTTGTCACACACTGTTCTGAATAATATATGCGG-----797
 QY 174 GlyGluAlaThrGlyProAspLeuGluGlyLysAspIleGlnThrGlyPheAlaGlyPro 193
 DB 796 -----TCTTCGATCTTCGCGATGAAGAGTTGAACT--GTTTTCACACA 752
 QY 194 SerGluAlaGluSerThrHisLeuAsp---ThrLysLysPro-----206

DB 751 TCTGAAGCCCAAGATACCATATAACCAAAAGCCCTCGTACTCGAATCTTGA 692
 QY 207 -----GlyTyrAsnGluIleProGluArgGluGluAsnGly 218
 DB 691 TTTCATTTCTCATCTACACTGTGTCACCCACAAAGACACTCCACAGAAAGAGGG 632
 QY 219 GlyAsnThrIleGlyThrArgAspGluThrAlaLysGlu-AlaAspAlaValAspValSe 238
 DB 631 CCCCACTCCATCTCCAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 572
 QY 238 rLeuValGlu-----GlySerAsnIleMetGlySerThrAsnPh 252
 DB 571 TCTCTTAAGAAATGAGTACACCGAGCCGGAATAATCCGTTCTCACTCCCA 517
 QY 252 eLysGluLeuProGluArgGluGlyAsnArgValAspAlaGlySerGlnAsnAla---Hi 271
 DB 516 -AAAGCATATTGCAAGAGAAAGAGGAGCAATCTCAGTCTCGTCCAAAAAGGATACT 458
 QY 271 sGlnGlyLysValGluPheHisTyrProAlaProSerLysGluLysArgLysGlu 291
 DB 457 ACTAGGAAAGCAGAGA-TCTGAATCAGTGTCCCAAGAGAAAGAAAGAAAGAAAG 399
 QY 291 Y-----SerSerAspAlaAlaGluSerThrAsnTyrAs 302
 DB 398 CAAAGATCTCCACCAAGTCAAGATTAATCTTCCAGAGAGAAATCCAGGTCCACAG 339
 QY 302 ngIuIleProLysAsnGlyLysGlySerThrArgLysGlyValAspHisSerAsnArgAs 322
 DB 338 CAGAGAACGAGAAAGATGATGATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 279
 QY 322 ngAlaIleThrLeuAsnGluLysGlnArgPheProSerLysGlyLysSerGlnGlyLeuPr 342
 DB 278 GTGGCTAGGTCACAGATCTATCTAGTCCCTCAAGATGAGAAC-----230
 QY 342 olleProSerArgLysLeuAspAsnGluIleLysAsnGluMetAspSerPheAsnGlyPr 362
 DB 229 -----AAAGTAAGATTCATCATTT-----209
 QY 362 oSerHisGluAsnIleIleThrHisGlyArgLysTyrHisTyrValProHisArgGlnAs 382
 DB 208 -----GCTAGA-----203
 QY 382 nasnSerThrArgAsnLysGlyMetProGlnGlyLysGlySerThrPoly-----398
 DB 202 -ATTGACAGAGATTAATCTCTCCCGGTGGAAGAGGAGGCAATATGATGTTG 147
 QY 399 -ArgGlnProHisSerAsnArgArgPheSerSerArgArgAspAspSerSerGlySe 418
 DB 146 GAGATGTCACAGAGAAATGATCGGTAC-----AGAAAGATGACCCAGAGAAACA 96
 QY 418 rSerAspSerGlySerSerSerGluSerAsp 428
 DB 95 GAATGAAAAATACAGAAAAAGAAAAAATGAC 65
 RESULT 13
 US-09-811-045A-2
 Sequence 2, Application US/09811045A
 Patent No. US20020035080A1
 GENERAL INFORMATION:
 APPLICANT: Scott, Robert E.
 TITLE OF INVENTION: cDNA encoding P2P proteins and use of P2P cDNA-
 TITLE OF INVENTION: derived antibodies and antisense reagents of
 TITLE OF INVENTION: in determining the proliferative potential of
 TITLE OF INVENTION: normal, abnormal and cancer cells in animals
 FILE REFERENCE: D6386D
 CURRENT APPLICATION NUMBER: US/09/811,045A
 PRIOR FILING DATE: 2001-03-16
 PRIOR APPLICATION NUMBER: US 08/801,308
 NUMBER OF SEQ ID NOS: 4
 SEQ ID NO 2
 LENGTH: 5173

TYPE: DNA
ORGANISM: Unknown
FEATURE:
NAME/KEY: CDNA
OTHER INFORMATION: P2P CDNA
US-09-811-045A-2

Alignment Scores:

Score: 0.00848
Length: 5173
Percent Similarity: 125.50
Best Local Similarity: 33.95%
Conservative: 98
Query Match: 20.29%
Mismatch: 66
Indels: 163
Caps: 158
22

US-09-700-696b-2 (1-430) x US-09-811-045A-2 (1-5173)

```

QY 6 SerIleSerAnlySGlUaenThrHIsAnGlyLeuArGmetSerIleTyProLySer 25
DB 2503 TCAGTATCTGACAAAGCAAG-----AGGCAAAAGATAGCCAAAGTA 2547
QY 26 ThrGlyAnlySGlUaenThrHIsAnGlyLeuArGmetSerIleTyProLySer 45
DB 2548 AAAAGTACAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 2601
QY 46 GluTyGlyAlaIleuIleArGAsnMetGlnHIsIleMetGlyProValThrAla 65
DB 2602 -----AATGTTTAAACCTTCAAGGACCTCAAGAAAG 2637
QY 66 IleLeuLeuLeuGlyGluGluAnlySGlUaenThrProArGAsn----- 80
DB 2638 GTA-----GATGAGACCTGAAAGATCTCTCGGTCTGAGCCGCTCAAA 2685
QY 81 -----ValLeuAnIleIleProAlaSerMetAsnTyR--- 91
DB 2686 AAAGCCAAAGAGAGGCTCAAAAGATGCTGATAAACTCTCGCTCTCTGAGAA 2745
QY 92 -----AlaLySaIaHIsSerLySaAspLySylsTyProGln 103
DB 2746 GATGAGAGGCTCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 2805
QY 104 ArgAspSerGln-----AlaGln 109
DB 2806 AGGCAAAAGCCAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 2865
QY 110 LysSer-----ProValLySerLySerThrHIsArgIleGlnHIsAnIleAspTyLeu 128
DB 2866 AAAAGCCAGCCAGTAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 2922
QY 129 -----LysHIsLeuSerLySylsTySylsIleProSerAspPheGlu 142
DB 2923 ATAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 2982
QY 143 GlySerGlyTyThrAspLeuGlnGluArgGlyAspAsnSylsIleSerPro----- 159
DB 2983 TCGCTTCAATGAATACTCTAAAGTAGAAGCAAGCAAGCAAGCAAGCAAG 3042
QY 160 -----PheSerGlyAspGlyGlnProPheLySaIleProGlyLySylsGluAlaThr 177
DB 3043 CGGAAATGAAAGCTGATGTTGAAAGCTGAAAGCAAGCAAGCAAGCAAGCAAG 3102
QY 178 GlyProAspLeuGlyGlyAspIle-----GlnThrGlyPheAlaGlyPro 193
DB 3103 TCATCAACTCTCCAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 3150
QY 194 SerGluAlaGluSerThrHIsLeuAspThrLySylsProGlyTyArgGlnIleProGlu 213
DB 3151 -----AAAAAATTGGA----- 3162
QY 214 ArgGluGluAnlySGlUaenThrHIsAnGlyLeuArGmetSerIleTyProLySer 233
DB 3163 -----AATGCAAAATGATCTACTACAAAGAAAGCAAGCAAGCAAGCAAG 3204

```

```

QY 234 AlaValAspValSerLeuValGluGlySerAsnAspIleMetGlySerThrAsnPhelys 253
DB 3204 ----- 3204
QY 254 GluLeuProGlyArgGlyGluGlyAsnArgValAspAlaGlySerGlnAsnAlaHIsGlnGly 273
DB 3205 -----TTGGAGTCAACCTTCAAGCAAGCAAGCAAGCAAGCAAG 3237
QY 274 LysValGluPheHIsTyRProProAlaProSerLySylsArgLySylsGlySer 293
DB 3238 AAAGTCAAG-----GGAAAGCCAAAGCAAGCAAGCAAGCAAGCAAG 3276
QY 294 AspAlaAlaGluSerThr-----AsnTyArgGlnIleProLySylsGlySer 311
DB 3277 GAAGGCTCAGCTCCAGCTTGTGATTCACCACTACAGTCAAGTCAAGTCAAGTCAAG 3336
QY 312 ThrArgLyGlyValAspHIsSerAsnArgGlnAlaThrLeuAnGlyGlnAlaArg 331
DB 3337 GTGAGCAATCTGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 3396
QY 332 Phe-----ProSerLySylsSerGlnGlyLeuProIleProSer 345
DB 3397 TATAATATGATTAACACAGCTCTGCTGATGATGATTAATTAATGATCAAGTCTCTCAG 3456
QY 346 ArgGlyLeuAspAsnGlu-----IleLySaNGluMetAspSerPheAsnGlyProSer 363
DB 3457 TCCAAATGGATTAAGATGACTTGTAGTCGAAAGCAAGATGTTAA-ACCACACAAAC 3515
QY 364 His-----GluAnIleIleThrHIsGlyArgLySylsTyRAlaProHIsArg 380
DB 3516 TATACAGAGTACGAAACCATCGAGATTAATTAATTAATTAATTAATTAATTAATTAAT 3574
QY 381 GlnAsnAsnSerThrArgAsnLySylsMetProGlnGlyLySylsSerTyRArgGln 400
DB 3575 CTACGGCTAGTACA-----CCGAGAAAGCAAGCAAGCAAGCAAGCAAG 3610
QY 401 ProHIsSerAsnArgArgPheSerSerArgArgArgAspSerSerGluSerAsp 420
DB 3611 CCGAAGAACTGCAAGCAAGCTTC-----CCAAGGAGCGAGCCAGCGTCAAGTCAAGCAGCAG 3676
QY 421 SerGlySer 423
DB 3668 TCAGGAGCT 3676

```

RESULT 14

US-10-149-110-6
Sequence 6, Application US/10149110
Publication No. US20030008330A1
GENERAL INFORMATION:
APPLICANT: Ade, Akhilo
APPLICANT: Eml, No. US20030008330A1uhiko
TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES, REMEDIES
FILE REFERENCE: 14114-003051
CURRENT APPLICATION NUMBER: US/10/149,110
PRIOR FILING DATE: 2002-06-06
PRIOR APPLICATION NUMBER: PCT/JP00/08813
PRIOR FILING DATE: 2000-12-13
NUMBER OF SEQ ID NOS: 10
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6
LENGTH: 2567
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(1122)
US-10-149-110-6

Alignment Scores:

Pred. No.: 0.00379 Length: 2567
 Score: 125.00 Matches: 81
 Percent Similarity: 31.49% Conservative: 44
 Best Local Similarity: 20.40% Mismatches: 146
 Query Match: 5.48% Indels: 126
 DB: 9 Gaps: 14

US-09-700-696B-2 (1-430) x US-10-149-110-6 (1-2567)

QY 78 ProArgAsnValLeuAsnIleLeuProAlaSerMetAsnTyrAlaLysAlaHisSerLys 97
 121 CCCAAATGTTAAACAGCCCTGCTCTCAGAGATGATCATCTACTAGAGAGACAAA 180
 QY 98 AspLysLysLysProGlnArgAspSerGlnAlaGlnLysSerProValLysSerLysSer 117
 181 AGTAAGAAAAAGGGGCGACAGAAAAACAACCTAAAGAGCAGCTTAAGAAAGATGAA 237
 QY 118 ThrHisArgLLeuGlnHisAsnIleAspTyrLeuLysHisLeuSerLysValLysLysIle 137
 237 -----
 QY 138 ProSerAspPheGlnGlySerGlyTyrThrAspLeuGlnGluArgLysAspAsnAspIle 157
 237 -----
 QY 158 SerProPheSerGlyAspGlyGlnProPheLysAspIleProGlyLysGlyGluAlaThr 177
 238 -----GAGGCCACAGAGAAAGATTAAGCCAAAGAAA----- 270
 QY 178 GlyProAspLeuGlnGlyLysAspIleGlnThrGlyPheAlaGlyProSerGlnAlaGlu 197
 271 GAGCCGATAAAAA-----GAGGGGAAAGAAAGAAAGTTGAA 306
 QY 198 SerThrHisLeuAspThrLysLysProGlyTyrAsnGlnIleProGluArgGluGluAsn 217
 307 TCAAAAGAAAAAATTTAGCTTAAACAGAGGCTTACTTCACTCCGATTCGAGAGAA 366
 QY 218 GlyGlyAsnThrIleGlyThrArgAsp-----GluThrAlaLys 230
 367 GGAGATGATCAAGAGGTGAAGAAAGAAAGAGGAGGAGAACTTTCACAGCTGCTCAG 426
 QY 231 GluAlaAspAlaValAspValSerLeuValGlnGlySerAsnAspIleMetGlySerThr 250
 427 AGAAGG-----AATATGCTGAAGGCCCAACATGAG----- 456
 QY 251 AsnPheLysGluLeuProGlyArg-----GluGlyAsnArgValAspAlaGlySer 267
 457 -----AAGAGAGCAGCAGATCGAAAAACGCAAGAGAGCAAAATGGAACCTGAGCAG 510
 QY 268 GlnAsnAlaHisGlnGlyLysValGluPheHisTyrProProAlaProSerLysGluLys 287
 511 CAAATAAAGAGTGAAGAGAA-----AAGCCAGAAAGTTAAGAAAGCTGAGAG 558
 QY 288 ArgLysGlnGlySerSerAspAla-----AlaGluSerThrAsnTyr 301
 559 AAGCGAAGAACATCATGATGCTCTCAGCTTCAAGAGTACATCATGATTAATAAATTCA 618
 QY 302 AsnGluLeuProLysAsnGlyLysGlySerThrArgLysGlyValAspHisSerAsnArg 321
 619 CTCAAATTGATTAATCTTGTATGTAACAGATGATGAGCCCTTGATGAACCTTGCTCA 678
 QY 322 AsnGlnAlaThrLeuAsnGlnLysGlnArgPheProSerLys----- 335
 679 CTTAGGTCAACATGCAACAGCTCAGAAACACACAGAGATGATTAAGTCACTGAAAAAA 738
 QY 336 -----GlyLysSerGlnGlyLeuProIleProSerArgLysLeuAspAsnGlu 351
 739 ATACGGCATTCAAAGTTAGTACAGATATCATGAGAAAGTCTACAAATGTTATTAACAG 798
 QY 352 IleLysAsnGluMetAspSerPheAsnGlyProSerHisGluAsnIleIleThrHisGly 371
 799 TTTAAGAAC-----ATGTTCTGTGTTGTAAGAGAGATTCGATGATCACCAGAGTG 849

QY 372 ArgLysTyrHisTyrValProHisArgGlnAsnAsnSerThrArgAsnLysGlyMetPro 391
 DB 850 CTGAATTAATCTCTTGTCTGACAAAGACACACATGAGGAAGCG---AATAAACCAAGAT 906
 QY 392 GlnGlyLys-----GlySerTyr----- 397
 DB 907 CAAGGAAAGAAAGGCCCAACAAAGCTAGAGAGAGAACAAACAGGCTCAAGACTCA 966
 QY 398 -----GlyArgGlnProHisSerAsnArgPheSerSer 409
 DB 967 AATGAGGATCTGATGCTCAAGATGTAATCATGACCAACATTAAGCGGAGACCAATGAA 1026
 QY 410 ArgArgArgAspAspSerSerGluSerSerAspSerGlySerSerGlu 426
 DB 1027 GACAGCAAGACACACATGAAAGCCAGACAGCAAAAAAGCCATTCAGTGA 1077

RESULT 15
 US-10-001-835-71
 Sequence 71, Application US/10001835
 Patent No. US20020160387A1
 GENERAL INFORMATION:
 APPLICANT: Salceda, Susana
 APPLICANT: Macina, Roberto
 APPLICANT: Recipon, Hervé
 APPLICANT: Cafferey, Robert
 APPLICANT: Sun, Yongming
 APPLICANT: Liu, Chenghua
 TITLE OF INVENTION: Compositions and Methods Relating to Ovary Specific Genes and
 FILE REFERENCE: DEX-0277
 CURRENT APPLICATION NUMBER: US/10/001,835
 CURRENT FILING DATE: 2001-11-20
 PRIOR APPLICATION NUMBER: 60/249,997
 PRIOR FILING DATE: 2000-11-20
 NUMBER OF SEQ ID NOS: 228
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 71
 LENGTH: 1374
 TYPE: DNA
 ORGANISM: Homo sapien
 US-10-001-835-71

Alignment Scores:
 Pred. No.: 0.00236 Length: 1374
 Score: 123.50 Matches: 78
 Percent Similarity: 33.86% Conservative: 51
 Best Local Similarity: 20.47% Mismatches: 169
 Query Match: 5.42% Indels: 83
 DB: 9 Gaps: 14

US-09-700-696B-2 (1-430) x US-10-001-835-71 (1-1374)

QY 93 LysAlaHisSerLysAspLysLysLysProGlnArgAspSerGlnAlaGlnLysSerPro 112
 DB 351 AAGAGAAAGAAAGACACAAACGAGACCGCGCAGCAACAAAGCAGCCGAGC----- 404
 QY 113 ValLysSerLysSerThrHisArgIleGlnHisAsnIleAspTyrLeuLysHisLeuSer 132
 DB 405 ---ACAGCGCAGAGCAGCAGAGCAGAGAGAGAGACCAAGCCAGC----- 443
 QY 133 LysValLysLysIleProSerAspPheGlnGlySerGlyTyrThrAspLeuGlnGluArg 152
 DB 444 GGCAGCGACAGAGCACCCACACAAAGAGAGGCGGATGCGGCGCGAGCAACAGAGAG 503
 QY 153 GlyAspAsnAspIleSerProPheSerGlyAsp-----GlyGln 165
 DB 504 GGTGATAGATATGATGAGGAGAGAGATGAGTTAGAGAGGTGATGCTATTAGAGAG 563
 QY 166 ProPheLysAspIleProGlyLysGlyGluAlaThr----- 177
 DB 564 CCCAAAGCAGAGAGAAAGAAATGAGAGAGAGAGCGACCGAGGATAGAGAGAGAGAGAG 623
 QY 178 GlyProAspLeuGlnGlyLysAspIleGlnThrGlyPheAlaGlyProSerGlnAlaGlu 197

```
Db 624 CCGAGGAC--GAGGACAGGAGCGCGAAGGAGAGGCGCAAGCGCGCCAGG 680
Qy 198 SerThrHisLeuAspThrLysLysProGlyTyrAsnGluLeuProGluArgGluAsn 217
Db 681 ---ACAGGCGCGAGCGCGAAGAGCGG-----GAAGAGCGACAGAGAGCG 722
Qy 218 GlyGlyAsnThrIleGlyThrArgAspGluThrAlaLysGluAlaAspAlaVal 237
Db 723 GAGGCGACAGAAAGAGCGAAGCGACAGAAAGAGACCGCGCGAGCAAGGAGAGCA 782
Qy 238 SerLeuValGluLysSerAsnAspIleMetGlySerThrAsnPheLysGluLeuProGly 257
Db 783 CCGAGAGAGAGAGGCGCGAAGAGAGAGGAGCGAGCGCGAGCGCGCGCGCGAGGC 842
Qy 258 ArgGluGlyAsnArgValAspAlaGly-----SerGluAsnAlaHisGlnGly 273
Db 843 CAGGAGACCGAAGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 902
Qy 274 LysValGluPheHisTyrProProAlaProSerLysGluLysArgLysGluLysSer 293
Db 903 GCCACGCCAG-----CGAGACGAGAGAGAGAG 929
Qy 294 AspAlaAlaGluSerThrAsnTyrAsnGluLeuProLysAsnGlyLysGlySerThrArg 313
Db 930 AGCGCAGAGCAGACCGCGCA-----GCGCGCGCGCGCGAGG 965
Qy 314 LysGlyValAsp-----HisSerAsnArgAsnGlnAlaThrLeu 965
Db 966 AGACCGAGCGAGCGCGCGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1025
Qy 327 AsnGluLysGlnArgPheProSerLysGlyLysSerGlnGlyLeuProIleProSerArg 346
Db 1026 CAGCGGAGAGCGCGCGCGCGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1073
Qy 347 GlyLeuAspAsnGluIleLysAsnGluMetLysAspSerPheAsnGlyProSerHisGluAsn 366
Db 1074 CCGAGGACGACAGAGAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1133
Qy 367 IleIleThrHisGlyArgLysTyrHisTyrValProHisArgGlnAsnAsnSerThrArg 386
Db 1134 GCCGCGAGAGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1193
Qy 387 -AsnLysGlyMetProGlnGlyLysGlySerTrpGlyArgGlnProHisSerAsn----- 404
Db 1194 AAGACCGAAGCGCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1253
Qy 405 -----ArgArgPheSerSerArg-ArgArg 413
Db 1254 CCGAGGACCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1313
Qy 413 AspAspSerSerGluSer-----SerAspSerGlySerSerSerGluSerAspGly 429
Db 1314 AAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1372
```

Search completed: April 17, 2003, 05:27:34
Job time : 173 secs

